

Parametric Inference for Nonsynchronously Observed Diffusion Processes in the Presence of Market Microstructure Noise

Teppei Ogihara

The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan
 School of Multidisciplinary Sciences, SOKENDAI (The Graduate University for Advanced Studies),
 Shonan Village, Hayama, Kanagawa 240-0193, Japan

Abstract. We study parametric inference for diffusion processes when observations occur nonsynchronously and are contaminated by market microstructure noise. We construct a quasi-likelihood function and study asymptotic mixed normality of maximum-likelihood- and Bayes-type estimators based on it. We also prove the local asymptotic normality of the model and asymptotic efficiency of our estimator when the diffusion coefficients are constant and noise follows a normal distribution. We conjecture that our estimator is asymptotically efficient even when the latent process is a general diffusion process. An estimator for the quadratic covariation of the latent process is also constructed. Some numerical examples show that this estimator performs better compared to existing estimators of the quadratic covariation.

Keywords. asymptotic efficiency, Bayes-type estimation, diffusion processes, local asymptotic normality, non-synchronous observations, parametric estimation, maximum-likelihood-type estimation, market microstructure noise

1 Introduction

Analysis of volatility and covariation is one of the most important subjects in the study of risk management of financial assets. Studies of high-frequency financial data are increasingly significant as high-frequency financial data become increasingly available and computing technology develops. While realized volatility has been studied as a consistent estimator of integrated volatility at high-frequency limits, estimators of covariation of two securities are also important. The realized covariance, a natural extension of the realized volatility, is a consistent estimator of integrated covariation in ideal settings.

However, there are two significant problems in empirical analysis, one of which is the existence of observation noise. When we model stock price data by a continuous stochastic process, we should assume that the observations are contaminated by additional noise as a way to explain empirical evidence. Consistent estimators of volatility under the presence of microstructure noise are investigated—for example, in Zhang, Mykland, and Aït-Sahalia [32], Barndorff-Nielsen et al. [3], and Podolskij and Vetter [27]—by using various data-averaging or resampling methods to reduce the influence of noise. The other significant problem is that of nonsynchronous observation, namely, that we observe prices of different securities at different time points. The realized covariance has serious bias under models of nonsynchronous observations, though we can calculate the estimator by using some simple ‘*synchronization*’ methods such as linear interpolation or the ‘previous tick’ methods. Hayashi and Yoshida [15, 16, 17] and Malliavin and Mancino [23, 24] independently constructed consistent estimators for statistical models of diffusion processes with nonsynchronous observations. There are also studies of covariation estimation under the simultaneous presence of microstructure noise and nonsynchronous observations. We refer interested readers to Barndorff-Nielsen et al. [4] for a kernel based method; Christensen, Kinnebrock, and Podolskij [7], Christensen, Podolskij, and Vetter [8] for a pre-averaged Hayashi–Yoshida estimator; Aït-Sahalia, Fan, and Xiu [2] for a method using the maximum likelihood estimator of a model with constant diffusion coefficients; and Bibinger et al. [5] for a technique employing the local method of moments.

While the above studies concern estimators under non- or semi-parametric settings, there are also studies about parametric inference of diffusion processes with high-frequency observations. Genon-Catalot and Ja-

cod [11] constructed quasi-likelihood function and studied an estimator that maximizes it. Gloter and Jacod [13] studied an estimator based on a quasi-likelihood function with noisy observations. Ogihara and Yoshida [26] studied a maximum-likelihood-type estimator and a Bayes-type estimator on nonsynchronous observations without market microstructure noise.

One advantage of maximum-likelihood- and Bayes-type estimators is that they are asymptotically efficient in many models. If a statistical model has the local asymptotic mixed normality (LAMN) property, then the results in Jeganathan [21, 22] ensure that asymptotic variance of estimators cannot be smaller than a certain lower bound. When some estimator attains this bound, it is called asymptotically efficient. For parametric estimation of diffusion processes on fixed intervals, Gobet [14] proved the LAMN property of the statistical model having equidistant observations, and an estimator in [11] is asymptotically efficient. Ogihara [25] proved the LAMN property and asymptotic efficiency of estimators for the setting of [26]. Gloter and Jacod [12] proved the local asymptotic normality (LAN) property for a statistical model with market microstructure noise when diffusion coefficients are deterministic, and the estimator by Gloter and Jacod [13] is asymptotically efficient. There are few studies about the efficiency of estimators that assume the presence of market microstructure noise and nonsynchronous observations. One exception is Bibinger et al. [5], who showed a lower bound of asymptotic variance of estimators in semi-parametric Cramér-Rao sense. We need the LAN or LAMN property of the statistical model to obtain asymptotic efficiency of a parametric model. To the best of our knowledge, this has not been studied for statistical models of noisy, nonsynchronous observations.

This paper examines consistency and asymptotic mixed normality of a maximum-likelihood-type estimator and a Bayes-type estimator based on a quasi-likelihood function, under the simultaneous presence of market microstructure noise and nonsynchronous observations. We also study the LAN property of this model when diffusion coefficients are constants, as well as the asymptotic efficiency of our estimators. We expect that our estimators are asymptotically efficient in the general cases. However, it is further difficult to obtain LAMN properties for models of general diffusions. This does not seem to have been obtained even for noisy, equidistant observations, and is left as future work. We will see by simulation that sample variance of the estimation error of our estimator is better than that of existing estimators for some examples in Section 3. These results ensure that our estimator not only is the theoretical best for asymptotic behavior, but also works well in practical finite samplings.

Our study has several advantages in addition to the above arguments regarding asymptotic efficiency.

- i) Our model also allows observation noise that follows a non-Gaussian distribution. We use a quasi-likelihood function for Gaussian noise, but our method is robust enough to allow misspecification of the noise distribution.
- ii) Since we obtain the results regarding asymptotic behaviors of the quasi-likelihood function as a byproduct, many applications become available from the theory of maximum-likelihood-type estimation. For example, we can construct a theory of the likelihood ratio test and one-step estimators as an immediate application. Further, the theory of information criteria is expected to follow from our results of quasi-likelihood functions.
- iii) Our settings contain random sampling schemes where the maximum length of observation intervals is not bounded by any constant multiplication of the minimum length. This is the case for some significant random sampling schemes, such as samplings based on Poisson or Cox processes. Our model encompasses such natural sampling schemes.

To obtain asymptotic mixed normality of our estimator, we investigate asymptotic behaviors of a quasi-likelihood function of noisy, nonsynchronous observations. To this end, we need to specify the limit of some matrix trace related to a ratio of covariance matrices for two different values of parameters, as appearing in (5.1). The inverse of the covariance matrix of observation noise has nontrivial off-diagonal elements, and so the inverse of the covariance matrix of observations is far from a diagonal matrix. This phenomenon is essentially different from the case of *synchronous* observations without noise (where the covariance matrix of observations is diagonal), and the case of *nonsynchronous* observations without noise (where the inverse of the covariance matrix is not a diagonal matrix but is ‘close’ to being one).

In a model of noisy, *synchronous* observations, the covariance matrix of a latent process is asymptotically equivalent to a unit matrix of the appropriate size, and is therefore simultaneously diagonalizable with the noise covariance. Gloter and Jacod [12, 13] used these facts and closed expressions for the eigenvalues of the noise

covariance to identify the limit of the quasi-likelihood function, but we cannot apply their idea because our sampling scheme is irregular and so not well approximated by a unit matrix. Further, the sizes of the covariance matrices are different for different components of the process, which follows from nonsynchronousness. In this paper, we deduce an asymptotically equivalent transform of the trace of the ratio of covariance matrices. This transform changes sizes of matrices and matrix elements into local averages, and arises from specific properties of the noise covariance matrix. We will see these results in Sections 4 and 5.

The remainder of this paper is organized as follows. In Section 2, we describe our detailed settings and main results. We propose a quasi-likelihood function for models with noisy, nonsynchronous observations, and construct a maximum-likelihood-type estimator based on it. We introduce asymptotic mixed normality of our estimator and results about asymptotic efficiency in Section 2.2. Section 2.3 contains results about the LAN property of our model and the asymptotic efficiency of our estimator, and Section 2.4 is devoted to results about Bayes-type estimators and convergence of moments of estimators. Polynomial-type large deviation inequalities, introduced in Yoshida [30, 31], are key to deducing these results. In Section 3 we will examine simulation results of our estimator for a simple example where the latent process is a Wiener process. We also construct an estimator of the quadratic covariation and compare the performance of our estimator with that of other estimators. The remaining sections are devoted to a proof of the main results. Section 4 introduces an asymptotically equivalent expression of the quasi-likelihood function. This expression is useful for deducing asymptotic properties of the quasi-likelihood function in Section 5. We also need some results on identifiability of the model to obtain consistency of the maximum-likelihood-type estimator. These are discussed in Section 6. Section 7 shows asymptotic mixed normality of our estimator. The LAN property of the model for constant diffusion coefficients is obtained in Section 8. Section 9 contains a proof of results regarding the Bayes-type estimator and the convergence of moments of estimators.

2 Main results

2.1 Settings and construction of the estimator

Let $(\Omega^{(0)}, \mathcal{F}^{(0)}, P^{(0)})$ be a probability space with a filtration $\mathbf{F}^{(0)} = \{\mathcal{F}_t^{(0)}\}_{0 \leq t \leq T}$. We consider a two-dimensional $\mathbf{F}^{(0)}$ -adapted process $Y = \{Y_t\}_{0 \leq t \leq T}$ satisfying the stochastic integral equation:

$$Y_t = Y_0 + \int_0^t \mu_s ds + \int_0^t b(s, X_s, \sigma_*) dW_s, \quad t \in [0, T], \quad (2.1)$$

where $\{W_t\}_{0 \leq t \leq T}$ is a d_1 -dimensional standard $\mathbf{F}^{(0)}$ -Wiener process, $b = (b^{ij})_{1 \leq i \leq 2, 1 \leq j \leq d_1}$ is a Borel function, $\mu = \{\mu_t\}_{0 \leq t \leq T}$ is a locally bounded $\mathbf{F}^{(0)}$ -adapted process with values in \mathbb{R}^2 , and $X = \{X_t\}_{0 \leq t \leq T}$ is a continuous $\mathbf{F}^{(0)}$ -adapted processes with values in O , an open subset of \mathbb{R}^{d_2} with $d_2 \in \mathbb{N}$. We consider market microstructure noise $\{\epsilon_i^{n,k}\}_{n \in \mathbb{N}, i \in \mathbb{Z}_+, k=1,2}$ as an independent sequence of random variables on another probability space $(\Omega^{(1)}, \mathcal{F}^{(1)}, P^{(1)})$. We assume that $\mathcal{F}^{(1)} = \mathfrak{B}((\epsilon_i^{n,k})_{n,k,i})$ and that the distribution of $\epsilon_j^{n,k}$ does not depend on j , where $\mathfrak{B}(S)$ denotes the minimal σ -field such that any element of S is $\mathfrak{B}(S)$ -measurable for a set S of random variables. We use the same notation $\mathfrak{B}(S)$ for a similarly defined σ -field for a set S of measurable sets. We consider a product probability space (Ω, \mathcal{F}, P) , where $\Omega = \Omega^{(0)} \times \Omega^{(1)}$, $\mathcal{F} = \mathcal{F}^{(0)} \otimes \mathcal{F}^{(1)}$, and $P = P^{(0)} \otimes P^{(1)}$.

We assume that the observations of processes occur in a nonsynchronous manner and are contaminated by market microstructure noise, that is, we observe the vectors $\{\tilde{Y}_i^k\}_{0 \leq i \leq \mathbf{J}_{k,n}, k=1,2}$ and $\{\tilde{X}_j^k\}_{0 \leq j \leq \mathbf{J}'_{k,n}, 1 \leq k \leq d_2}$, where $\{S_i^{n,k}\}_{i=0}^{\mathbf{J}_{k,n}}$ and $\{T_j^{n,k}\}_{j=0}^{\mathbf{J}'_{k,n}}$ are random times in $(\Omega^{(0)}, \mathcal{F}^{(0)})$, $\{\eta_j^{n,k}\}_{j \in \mathbb{Z}_+, 1 \leq k \leq d_2}$ is a random sequence on (Ω, \mathcal{F}) , and

$$\tilde{Y}_i^k = Y_{S_i^{n,k}}^k + \epsilon_i^{n,k}, \quad \tilde{X}_j^k = X_{T_j^{n,k}}^k + \eta_j^{n,k}. \quad (2.2)$$

Our goal is to estimate the true value σ_* of the parameter from nonsynchronous, noisy observations $\{S_i^{n,k}\}_{0 \leq i \leq \mathbf{J}_{k,n}, k=1,2}$, $\{T_j^{n,k}\}_{0 \leq j \leq \mathbf{J}'_{k,n}, 1 \leq k \leq d_2}$, $\{\tilde{Y}_i^k\}_{0 \leq i \leq \mathbf{J}_{k,n}, k=1,2}$, and $\{\tilde{X}_j^k\}_{0 \leq j \leq \mathbf{J}'_{k,n}, 1 \leq k \leq d_2}$.

By setting $d_2 = 2$, $X_t \equiv Y_t$, $\mu_t = \mu(t, Y_t)$, $S_i^{n,k} \equiv T_j^{n,k}$, and $\eta_j^{n,k} \equiv \epsilon_i^{n,k}$, our model contains the case where the latent process Y is a diffusion process satisfying a stochastic differential equation

$$dY_t = \mu(t, Y_t)dt + b(t, Y_t, \sigma_*)dW_t, \quad t \in [0, T], \quad (2.3)$$

and Y is observed in a nonsynchronous manner with noise. This model is of particular interest, but our results are also be applied to more general models (2.1).

Remark 2.1. *Stochastic volatility models are significant models for modeling stock prices. Unfortunately, our settings are not applied to hidden Markov models including stochastic volatility models because we require (possibly noisy) observations of process X . However, we hope that our results give an essential idea to deal with noisy, nonsynchronous observations, and therefore we can construct an estimator for stochastic volatility models by replacing our quasi-likelihood function. We have left it for future works.*

For a vector $x = (x_1, \dots, x_k)$, we denote $\partial_x^l = (\frac{\partial^l}{\partial x_{i_1} \dots \partial x_{i_l}})_{i_1, \dots, i_l=1}^k$. We assume the true value σ_* of the parameter is contained in a bounded open set $\Lambda \subset \mathbb{R}^d$ that satisfies Sobolev's inequality; that is, for any $p > d$, there exists $C > 0$ such that $\sup_{\sigma \in \Lambda} |u(x)| \leq C \sum_{k=0,1} (\int_{\Lambda} |\partial_x^k u(x)|_p dx)^{1/p}$ for any $u \in C^1(\Lambda)$. This is the case when Λ has a Lipschitz boundary. See Adams and Fournier [1] for more details.

Let $\Pi_n = (\{S_i^{n,k}\}_{n,k,i}, \{T_j^{n,k}\}_{n,k,j})$ and $\{\mathcal{G}_t\}_{0 \leq t \leq T}$ be a filtration of (Ω, \mathcal{F}, P) given by

$$\mathcal{G}_t = \mathcal{F}_t^{(0)} \bigvee \mathfrak{B}(\{\Pi_n\}_n) \bigvee \mathfrak{B}(A \cap \{S_i^{n,k} \leq t\}; A \in \mathfrak{B}(\epsilon_i^{n,k}), m \in \mathbb{N}, k \in \{1, 2\}, i \in \mathbb{Z}_+, n \in \mathbb{N}),$$

where $\mathcal{H}_1 \bigvee \mathcal{H}_2$ denotes the minimal σ -field which contains σ -fields \mathcal{H}_1 and \mathcal{H}_2 . We assume that $(X_t, Y_t, W_t, \mu_t)_t$ and $(\{S_i^{n,k}\}_{n,k,i}, \{T_j^{n,k}\}_{n,k,j})$ are independent. Moreover, we assume that there exist positive constants $v_{1,*}$ and $v_{2,*}$ such that $\eta_j^{n,k} 1_{\{T_j^{n,k} \leq t\}}$ is \mathcal{G}_t -measurable,

$$E[\epsilon_i^{n,k} 1_{\{S_i^{n,k} > t\}} | \mathcal{G}_t] = 0, \quad E[\epsilon_i^{n,k} \epsilon_{i'}^{n,k'} 1_{\{S_i^{n,k} \wedge S_{i'}^{n,k'} > t\}} | \mathcal{G}_t] = v_{k,*} \delta_{ii'} \delta_{kk'}$$

for any n, k, k', i, i', j, t , where δ_{ij} is Kronecker's delta and $E_{\Pi}[\mathbf{X}] = E[X | \{\Pi_n\}_n]$ for a random variable \mathbf{X} . We also assume that the distribution of Y_0 does not depend on σ_* , $v_{1,*}$, nor $v_{2,*}$.

Now we construct the quasi-likelihood function. We apply the idea of Gloter and Jacod [13] to our construction of a quasi-likelihood function; that is, we divide the whole observation interval $[0, T]$ into equidistant subdivisions and construct quasi-likelihood functions for each interval as follows. Let $\{b_n\}_{n \in \mathbb{N}}$ and $\{k_n\}_{n \in \mathbb{N}}$ be sequences of positive numbers satisfying $b_n \geq 1$, $k_n \leq b_n$, $b_n \rightarrow \infty$, $k_n b_n^{-1/2-\epsilon} \rightarrow \infty$, and $k_n b_n^{-2/3+\epsilon} \rightarrow 0$ as $n \rightarrow \infty$ for some $\epsilon > 0$. We will assume in Condition [A2] a relation between b_n and our sampling scheme, which implies that b_n represents the order of observation frequency. Let $\ell_n = [b_n k_n^{-1}]$, $s_0 = 0$, $s_m = T[b_n k_n^{-1}]^{-1} m$, $b^k(t, x, \sigma) = (b^{kj}(t, x, \sigma))_{j=1}^{d_1}$, $K_0^k = -1$, and $K_m^k = \#\{i \in \mathbb{N}; S_i^{n,k} < s_m\}$ for $k \in \{1, 2\}$ and $1 \leq m \leq \ell_n$. Moreover, let $k_m^j = K_m^j - K_{m-1}^j - 1$, $\bar{k}_n = \max_{m,j} k_m^j$, $\underline{k}_n = \min_{m,j} k_m^j$, $J_m^k = \max\{1 \leq j \leq \mathbf{J}_{k,n}^j; T_j^{n,k} \leq s_{m-1}\}$, $I_{i,m}^k = [S_{i+K_{m-1}^k}^{n,k}, S_{i+1+K_{m-1}^k}^{n,k})$, $\tilde{Y}^k(I_{i,m}^k) = \tilde{Y}_{i+1+K_{m-1}^k}^k - \tilde{Y}_{i+K_{m-1}^k}^k$, $\hat{X}_m = (\#\{j; T_j^{n,k} \in [s_{m-1}, s_m)\})^{-1} \sum_{j; T_j^{n,k} \in [s_{m-1}, s_m)} \tilde{X}_j^k$, $1 \leq k \leq d_2$, and $b_m^j(\sigma) = b^j(s_{m-1}, \hat{X}_m, \sigma)$ for $1 \leq m \leq \ell_n$, $j \in \{1, 2\}$ and $1 \leq i \leq k_m^j$. Then we have the following approximations of conditional covariance of observations:

$$\begin{aligned} E[\tilde{Y}^k(I_{i,m}^k) \tilde{Y}^k(I_{i',m}^k) | \mathcal{G}_{s_{m-1}}] &\sim (|b_m^k|^2 |I_{i,m}^k| + 2v_k) \delta_{ii'} - v_1 1_{\{|i-i'|=1\}}, \\ E[\tilde{Y}^1(I_{i',m}^1) \tilde{Y}^2(I_{i'',m}^2) | \mathcal{G}_{s_{m-1}}] &\sim b_m^1 \cdot b_m^2 |I_{i',m}^1 \cap I_{i'',m}^2| \end{aligned} \quad (2.4)$$

for any intervals $I_{i,m}^k, I_{i',m}^k, I_{i'',m}^1, I_{i''',m}^2$.

Let \top denotes the transpose operator for matrices (and vectors), $M(l) = \{2\delta_{i_1, i_2} - \delta_{|i_1 - i_2|=1}\}_{i_1, i_2=1}^l$ for $l \in \mathbb{N}$, $M_{j,m} = M(k_m^j)$ for $1 \leq j \leq 2$. Based on the relation (2.4), we define a quasi-log-likelihood function $H_n(\sigma, v)$ by

$$H_n(\sigma, v) = -\frac{1}{2} \sum_{m=2}^{\ell_n} Z_m^\top S_m^{-1}(\sigma, v) Z_m - \frac{1}{2} \sum_{m=2}^{\ell_n} \log \det S_m(\sigma, v), \quad (2.5)$$

where $Z_m = ((\tilde{Y}^1(I_{i,m}^1))_{1 \leq i \leq k_m^1}, (\tilde{Y}^2(I_{i,m}^2))_{1 \leq i \leq k_m^2})^\top$ and

$$S_m(\sigma, v) = \begin{pmatrix} \{|b_m^1|^2 |I_{i,m}^1| \delta_{ii'}\}_{ii'} & \{b_m^1 \cdot b_m^2 |I_{i,m}^1 \cap I_{j,m}^2|\}_{ij} \\ \{b_m^1 \cdot b_m^2 |I_{i,m}^1 \cap I_{j,m}^2|\}_{ji} & \{|b_m^2|^2 |I_{j,m}^2| \delta_{jj'}\}_{jj'} \end{pmatrix} + \begin{pmatrix} v_1 M_{1,m} & 0 \\ 0 & v_2 M_{2,m} \end{pmatrix}. \quad (2.6)$$

Remark 2.2. Though such a local Gaussian quasi-log-likelihood function seems valid only when observation noise $\epsilon_i^{n,k}$ follows a Gaussian distribution, asymptotic properties of the maximum likelihood estimator are robust enough to allow non-Gaussian noise. We can use the same quasi-likelihood function for general noise.

Remark 2.3. We used subdivisions of $[0, T]$ for the construction of H_n because of technical issues related to deducing the limit of H_n . Since the diffusion coefficient b in S_m is fixed, matrix properties of $M_{j,m}$ introduced in Section 4.2 can be used to deduce the limit of H_n . On the other hand, such a construction of H_n also contributes to reducing the calculation time of the maximum-likelihood-type estimator because the size of S_m is $O(k_n)$ while the size of the covariance matrix of all observations is $O(b_n)$.

Remark 2.4. In [13], k_n is taken so that $n^{1/2}k_n^{-1} \rightarrow 0$ and $k_n n^{-3/4} \rightarrow 0$. Our rate $b_n^{2/3}$ for the upper bound of k_n is a little bit worse because of some technical issue (for equidistance observations, we have $b_n \equiv n$). When we investigate asymptotic behaviors of the maximum-likelihood-type estimator, we deal with some supremum estimates for the σ of quasi-likelihood ratios. Unlike the one-dimensional settings of [13], our multidimensional setting requires some properties to deal with the supremum. We use Sobolev's inequality here for this purpose. Then we need an additional moment estimate for quasi-likelihood ratios, which causes a worse rate of k_n . See the proofs of Lemmas 4.3 and 4.4 for details.

To construct the maximum-likelihood-type estimator $\hat{\sigma}_n$ for the parameter σ , we need estimators for the unknown noise variance $v_* = (v_{1,*}, v_{2,*})$. We assume the following condition.

[V] There exist estimators $\{\hat{v}_n\}_{n \in \mathbb{N}}$ of v_* such that $\hat{v}_n \geq 0$ almost surely and $\{b_n^{1/2}(\hat{v}_n - v_*)\}_{n \in \mathbb{N}}$ is tight.

For example, $\hat{v}_n = (\hat{v}_{n,k})_{k=1}^2$ with $\hat{v}_{n,k} = (2\mathbf{J}_{k,n})^{-1} \sum_i (\tilde{Y}_i^k - \tilde{Y}_{i-1}^k)^2$ satisfies [V] if $\{b_n \mathbf{J}_{k,n}^{-1}\}_n$ is tight for $k = 1, 2$, $\sup_{n,k,i} E[(\epsilon_i^{n,k})^4] < \infty$ and $\sup_{n,k,i \neq j} b_n^2 E[(\epsilon_i^{n,k})^2 - v_{k,*})(\epsilon_j^{n,k})^2 - v_{k,*}] < \infty$.

Let $\text{clos}(A)$ be the closure of a set A . A maximum-likelihood-type estimator $\hat{\sigma}_n$ is a random variable satisfying $H_n(\hat{\sigma}_n, \hat{v}_n) = \max_{\sigma \in \text{clos}(\Lambda)} H_n(\sigma, \hat{v}_n)$. We study asymptotic mixed normality and asymptotic efficiency of the estimator in the following subsections.

Remark 2.5. We can also construct a simultaneous maximum-likelihood-type estimator $(\bar{\sigma}_n, \bar{v}_n)$ satisfying $H_n(\bar{\sigma}_n, \bar{v}_n) = \max_{\sigma, v} H_n(\sigma, v)$. However, it is valid only when the observation noise $\epsilon_i^{n,k}$ follows a normal distribution. Our interest is on estimating the parameter σ of the latent process, and so the assumptions for observation noise should be reduced as much as possible. Therefore, the nonparametric estimator \hat{v}_n is more suitable for our purpose.

2.2 Asymptotic mixed normality of the maximum-likelihood-type estimator

In the rest of this section, we state our main theorems. Proofs of these results are left to Sections 4–9. In this subsection, we describe the asymptotic mixed normality of the maximum-likelihood-type estimator $\hat{\sigma}_n$.

We first describe assumptions for the theorem. Condition [A1] is a sequence of assumptions on the latent processes Y and X and observation noise $\epsilon_i^{n,k}$ and $\eta_j^{n,k}$. We denote by \mathcal{E}_l the unit matrix of size l .

- [A1] 1. For $0 \leq 2i + j \leq 4$ and $0 \leq k \leq 4$, the derivatives $\partial_t^i \partial_x^j \partial_\sigma^k b(t, x, \sigma)$ exist on $[0, T] \times O \times \Lambda$ and have continuous extensions on $[0, T] \times O \times \text{clos}(\Lambda)$.
2. $bb^\top(t, x, \sigma)$ is positive definite for $(t, x, \sigma) \in [0, T] \times O \times \text{clos}(\Lambda)$.
3. $\sup_{n,k,i} E[(\epsilon_i^{n,k})^q] < \infty$ for any $q > 0$.
4. μ_t is locally bounded (locally in time).
5. $\sup_n (\ell_n^{q/2} \max_{m,k} (\#\{j; T_j^{n,k} \in [s_{m-1}, s_m]\})^{-1} E_\Pi[\|\sum_{j; T_j^{n,k} \in [s_{m-1}, s_m]} \eta_j^{n,k}\|^q]) < \infty$ almost surely for any $q > 0$.
6. There exist progressively measurable processes $\{b_t^{(j)}\}_{0 \leq t \leq T, 0 \leq j \leq 1}$ and $\{\hat{b}_t^{(j)}\}_{0 \leq t \leq T, 0 \leq j \leq 1}$ such that $b_t^{(j)}$, $\hat{b}_t^{(j)}$, and $\sup_{u < s \leq t} (|b_s^{(j)} - b_u^{(j)}| \vee |\hat{b}_s^{(j)} - \hat{b}_u^{(j)}|)/|s - u|^{1/2}$ are locally bounded processes for $0 \leq j \leq 1$, and

$$X_t = X_0 + \int_0^t b_s^{(0)} ds + \int_0^t b_s^{(1)} dW_s, \quad b_t^{(1)} = b_0^{(1)} + \int_0^t \hat{b}_s^{(0)} ds + \int_0^t \hat{b}_s^{(1)} dW_s$$

for $t \in [0, T]$.

Condition [A1] captures somewhat standard assumptions and whether it holds can easily be verified in practical settings. Roughly speaking, point 5 of [A1] is satisfied if the summation of $\eta_j^{n,k}$ is of an order equivalent to the square root of the number of $\eta_j^{n,k}$. This is satisfied under certain independency, martingale conditions or mixing conditions of $\eta_j^{n,k}$. If $\{\eta_j^{n,k}\}_j$ is a sequence of independent and identically distributed values and the sequence has finite moments, then $E_\Pi[\sum_{j; T_j^{n,k} \in [s_{m-1}, s_m)} \eta_j^{n,k} |^q] = O_p(\#\{j; T_j^{n,k} \in [s_{m-1}, s_m)\}^{-q/2})$. Then, point 5 of [A1] is satisfied if sampling frequency of $\{T_j^{n,k}\}$ is of order b_n . Decomposition of X in point 6 of [A1] is used to deduce asymptotically equivalent representation of H_n where the diffusion coefficient $b(t, X_t, \sigma_*)$ is replaced by $b(s_{m-1}, X_{s_{m-1}}, \sigma_*)$. Detailed semimartingale decomposition is required to estimate the difference $b(t, X_t, \sigma_*) - b(s_{m-1}, X_{s_{m-1}}, \sigma_*)$.

In the following, we assume some conditions about our sampling scheme. For $\eta \in (0, 1/2)$, let \mathcal{S}_η be the set of all sequences $\{[s'_{n,l}, s''_{n,l}]\}_{n \in \mathbb{N}, 1 \leq l \leq L_n}$ of intervals on $[0, T]$ satisfying $\{L_n\}_n \subset \mathbb{N}$, $[s'_{n,l_1}, s''_{n,l_1}) \cap [s'_{n,l_2}, s''_{n,l_2}) = \emptyset$ for $n, l_1 \neq l_2$, $\inf_{n,l} (b_n^{1-\eta}(s''_{n,l} - s'_{n,l})) > 0$, and $\sup_{n,l} (b_n^{1-\eta}(s''_{n,l} - s'_{n,l})) < \infty$. Let $r_n = \max_{i,k,m} |I_{i,m}^k|$ and $\underline{r}_n = \min_{i,k,m} |I_{i,m}^k|$.

[A2] There exist $\eta \in (0, 1/2)$, $\dot{\eta} \in (0, 1]$ and positive-valued functions $\{a_t^j\}_{t \in [0, T], j=1,2}$ such that $\sup_{t \neq s} (|a_t^j - a_s^j|/|t - s|^{\dot{\eta}}) < \infty$ almost surely, $b_n^{-1/2} k_n (b_n^{-1} k_n)^{\dot{\eta}} \rightarrow 0$ and

$$k_n b_n^{-1/2} \max_{1 \leq l \leq L_n} \left| b_n^{-1} (s''_{n,l} - s'_{n,l})^{-1} \#\{i; [S_{i-1}^{n,j}, S_i^{n,j}) \subset (s'_{n,l}, s''_{n,l})\} - a_{s'_{n,l}}^j \right| \rightarrow^p 0 \quad (2.7)$$

as $n \rightarrow \infty$ for $j = 1, 2$ and $\{[s'_{n,l}, s''_{n,l}]\}_{1 \leq l \leq L_n, n \in \mathbb{N}} \in \mathcal{S}_\eta$. Moreover, $(r_n b_n^{1-\epsilon}) \vee (b_n^{1-\epsilon} \underline{r}_n^{-1}) \rightarrow^p 0$ for any $\epsilon > 0$.

In particular, Condition [A2] implies $b_n^{-1} \mathbf{J}_{j,m} \rightarrow^p \int_0^T a_t^j dt$ and $\max_m |T^{-1} k_n^{-1} k_m^j - a_{s_{m-1}}^j| \rightarrow^p 0$ as $n \rightarrow \infty$. Roughly speaking, [A2] shows the law of large numbers for sampling schemes in any local time intervals. In the proof of Lemma 5.2, we will see that some properties of $M_{j,m}$ enable us to replace $|I_k^j|$ in S_m by the local average in asymptotics. Then [A2] leads to the limit of H_n .

Example 2.1. Let $\{N_t^k\}_{t \geq 0}$ be an exponential α -mixing point process with stationary increments for $k = 1, 2$. Set $S_i^{m,k} = \inf\{t \geq 0; N_{b_n t}^k \geq i\}$. Then Rosenthal-type inequalities (Theorem 3 and Lemma 7 in Doukhan and Louhichi [10], or Theorem 4 in [26]) and a similar argument to the proof of Proposition 6 in [26] ensure [A2] with $a_t^j \equiv E[N_1^j]$ (constants).

Under the above conditions, we can show convergence of the quasi-likelihood ratio $H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n)$. The limit function is rather complicated, so we prepare some functions. Let $b_t = b(t, X_t, \sigma)$, $b_{t,*} = b(t, X_t, \sigma_*)$, $\tilde{a}_t^j = a_t^j / v_{j,*}$ for $j = 1, 2$, $\varphi(x, y) = \sqrt{x + \sqrt{x^2 - 4y}} + \sqrt{x - \sqrt{x^2 - 4y}}$ for $0 \leq 4y \leq x^2$, and

$$\begin{aligned} \mathcal{Y}_1(\sigma) = & \int_0^T \left\{ \frac{\sum_{j=1}^2 (|b_t^j|^2 - |b_{t,*}^j|^2) (|b_t^{3-j}|^2 \sqrt{\tilde{a}_t^1 \tilde{a}_t^2} + \tilde{a}_t^j \sqrt{\det(b_t b_t^\top)}) - 2(b_t^1 \cdot b_t^2 - b_{t,*}^1 \cdot b_{t,*}^2) b_t^1 \cdot b_t^2 \sqrt{\tilde{a}_t^1 \tilde{a}_t^2}}{2\sqrt{2} \sqrt{\det(b_t b_t^\top)} \varphi(\tilde{a}_t^1 |b_t^1|^2 + \tilde{a}_t^2 |b_t^2|^2, \tilde{a}_t^1 \tilde{a}_t^2 \det(b_t b_t^\top))} \right. \\ & \left. - \frac{\varphi(\tilde{a}_t^1 |b_t^1|^2 + \tilde{a}_t^2 |b_t^2|^2, \tilde{a}_t^1 \tilde{a}_t^2 \det(b_t b_t^\top)) - \varphi(\tilde{a}_t^1 |b_{t,*}^1|^2 + \tilde{a}_t^2 |b_{t,*}^2|^2, \tilde{a}_t^1 \tilde{a}_t^2 \det(b_{t,*} b_{t,*}^\top))}{2\sqrt{2}} \right\} dt. \end{aligned}$$

Proposition 2.1. Assume [A1], [A2] and [V]. Then $\sup_{\sigma \in \Lambda} |b_n^{-1/2} \partial_\sigma^k (H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n)) - \partial_\sigma^k \mathcal{Y}_1(\sigma)| \rightarrow^p 0$ as $n \rightarrow \infty$ for $0 \leq k \leq 3$.

To show consistency and asymptotic normality of $\hat{\sigma}_n$, the limit function $\mathcal{Y}_1(\sigma)$ of the quasi-likelihood ratio should have the unique maximum point at $\sigma = \sigma_*$. More precisely, we use the following as a kind of identifiability condition: $\inf_{\sigma \neq \sigma_*} (-\mathcal{Y}_1(\sigma)) / |\sigma - \sigma_*|^2 > 0$ almost surely. Though it is difficult to directly check this condition in general, we can check it under a more tractable sufficient condition. Let

$$\mathcal{Y}_0(\sigma) = -\frac{1}{2} \int_0^T \left\{ \text{tr}((b_t b_t^\top)^{-1} (b_{t,*} b_{t,*}^\top) - \mathcal{E}_2) + \log \frac{\det(b_t b_t^\top)}{\det(b_{t,*} b_{t,*}^\top)} \right\} dt.$$

Then \mathcal{Y}_0 is the probability limit $n^{-1/2} (H_n^0(\sigma) - H_n^0(\sigma_*))$, where H_n^0 represents a quasi-likelihood function for a statistical model of equidistant observations without noise.

[A3] $\inf_{\sigma \neq \sigma_*} ((-\mathcal{Y}_0(\sigma))/|\sigma - \sigma_*|^2) > 0$ almost surely.

We will show in Proposition 6.1 that [A3] is sufficient for the identifiability condition of our model. Moreover, the following condition is a simple sufficient condition for [A3] (see Remark 4 in Ogihara and Yoshida [26] for the details):

[A3'] $\inf_{\sigma_1 \neq \sigma_2} (|bb^\top(t, x, \sigma_1) - bb^\top(t, x, \sigma_2)|/|\sigma_1 - \sigma_2|) > 0$ for any $t \in [0, T]$ and $x \in O$.

We denote by $\rightarrow^{s-\mathcal{L}}$ the stable convergence of random variables. Let

$$\hat{\Gamma}_{1,n} = -b_n^{-1/2} \partial_\sigma^2 H_n(\hat{\sigma}_n, \hat{v}_n), \quad \Gamma_1 = -\partial_\sigma^2 \mathcal{Y}_1(\sigma_*). \quad (2.8)$$

Let \mathcal{N} be a d -dimensional random variable on some extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ of (Ω, \mathcal{F}, P) satisfying the condition that \mathcal{N} is independent of \mathcal{F} and \mathcal{N} follows the d -dimensional standard normal distribution. We denote the expectation with respect to \tilde{P} by the same notation E .

The following theorem is one of our main results.

Theorem 2.1. *Assume [A1]–[A3] and [V]. Then Γ_1 is positive definite almost surely and $b_n^{1/4}(\hat{\sigma}_n - \sigma_*) \rightarrow^{s-\mathcal{L}} \Gamma_1^{-1/2} \mathcal{N}$ as $n \rightarrow \infty$. Moreover, $\hat{\Gamma}_{1,n} \rightarrow^p \Gamma_1$, and therefore $b_n^{1/4} \hat{\Gamma}_{1,n}^{-1/2} 1_{\{\hat{\Gamma}_{1,n} \text{ is p.d.}\}}(\hat{\sigma}_n - \sigma_*) \rightarrow^{s-\mathcal{L}} \mathcal{N}$ as $n \rightarrow \infty$.*

Corollary 2.1. *Assume [A1], [A2], [A3'] and [V]. Then the results in Theorem 2.1 hold true.*

2.3 On the LAMN property and asymptotic efficiency of the estimator

In this subsection, we state some results on the so-called LAMN(LAN) property for our model and asymptotic efficiency of our estimator. We also comment on some further studies.

Throughout this subsection, we assume that $X_t \equiv Y_t$, $T_j^{n,k} \equiv S_i^{n,k}$, $\eta_j^{n,k} \equiv \epsilon_i^{n,k}$, $\mu_t \equiv 0$ and $Y_0 = \gamma$ for some known $\gamma \in \mathbb{R}^2$. Then the latent process Y is a diffusion process satisfying the stochastic differential equation (2.3) with $\mu \equiv 0$. Let $P_{\sigma'_*, v'_*, n}$ be the distribution of $((S_i^{n,k})_{k,i}, (\tilde{Y}_i^k)_{k,i})$ with true values (σ'_*, v'_*) of the parameters. We denote

$$\text{diag}(A, B) = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

for square matrices A and B . Let $\mathcal{Y}_2(v) = -\int_0^T \sum_{j=1}^2 a_t^j \{(v_{j,*}/v_j) - 1 + \log(v_j/v_{j,*})\} dt/2$,

$$\Gamma_2 = -\partial_v^2 \mathcal{Y}_2(v_*) \quad \text{and} \quad \Gamma = \text{diag}(\Gamma_1, \Gamma_2). \quad (2.9)$$

We adopt the following definition of the LAMN property from Jeganathan [22].

Definition 2.1. *Let $P_{\theta,n}$ be a probability measure on some measurable space $(\mathcal{X}_n, \mathcal{A}_n)$ for each $\theta \in \Theta$ and $n \in \mathbb{N}$, where Θ is a bounded open subset of \mathbb{R}^d . Then the family $\{P_{\theta,n}\}_{\theta,n}$ satisfies the local asymptotic mixed normality (LAMN) property at $\theta = \theta_*$ if there exist a sequence $\{\delta_n\}_{n \in \mathbb{N}}$ of $d \times d$ positive definite matrices, $d \times d$ symmetric random matrices Γ_n, Γ and d -dimensional random vectors $\mathcal{N}_n, \mathcal{N}$ such that Γ is positive definite a.s., $P_{\theta_*,n}[\Gamma_n \text{ is positive definite}] = 1$ ($n \in \mathbb{N}$), $\|\delta_n\| \rightarrow 0$, and*

$$\log \frac{dP_{\theta_* + \delta_n u, n}}{dP_{\theta_*, n}} - \left(u^\top \sqrt{\Gamma_n} \mathcal{N}_n - \frac{1}{2} u^\top \Gamma_n u \right) \rightarrow 0$$

in $P_{\theta_,n}$ -probability as $n \rightarrow \infty$ for any $u \in \mathbb{R}^d$. Moreover, \mathcal{N} follows the d -dimensional standard normal distribution, \mathcal{N} is independent of Γ and $\mathcal{L}(\mathcal{N}_n, \Gamma_n | P_{\theta_*, n}) \rightarrow \mathcal{L}(\mathcal{N}, \Gamma)$ as $n \rightarrow \infty$.*

If further the limit matrix Γ is non-random, we say $\{P_{\theta,n}\}_{\theta,n}$ has the local asymptotic normality (LAN) property.

To prove the LAMN property of our model, we assume the following additional condition.

[A1''] [A1] is satisfied, $\mu_t \equiv 0$, $b(t, x, \sigma)$ does not depend on (t, x) and $\epsilon_i^{n,k}$ follows a normal distribution for any n, k, i .

Theorem 2.2. Assume [A1''], [A2] and [A3]. Then the family of distributions $\{P_{\sigma_*, v_*, n}\}_{\sigma_*, v_*, n}$ has the LAN property with Γ in (2.9) and $\delta_n = \text{diag}(b_n^{-1/4}\mathcal{E}_d, b_n^{-1/2}\mathcal{E}_2)$.

Remark 2.6. Jeganathan [21] studied lower bounds of estimation errors for any estimator of parameters. They showed a version of Hájek's convolution theorem and that the optimal asymptotic variance of errors for regular estimators is Γ^{-1} , where Γ is in Definition 2.1. Therefore, Theorems 2.1 and 2.2 ensures that our estimator $\hat{\sigma}_n$ of the parameter σ is asymptotically efficient in this sense under the assumptions of both theorems.

Remark 2.7. The assumptions of Theorem 2.2 are rather strong conditions. We are also interested in the LAMN property in more general settings. In particular, we are interested in the case that $\mu_t = \mu(t, X_t)$ and μ and b are general functions with suitable conditions. However, we need further analysis using Malliavin calculus to deal with the LAMN property of general diffusion processes, as seen in Gobet [14] and Ogihara [25]. To the best of author's knowledge, such a result has not been obtained even for models with noisy, synchronous observations. We have left this for future works.

2.4 A Bayes-type estimator and convergence of moments of estimation errors

Polynomial-type large deviation theory by Yoshida [30, 31] enables us to address the asymptotic properties of a Bayes-type estimator and the convergence of moments of estimation errors, which is a stronger result than asymptotic mixed normality. Convergence of moments is useful when we investigate the theory of information criteria, minimax inequality and asymptotic expansion of estimators. See Uchida [28] for a theory of contrast-based information criteria for ergodic diffusion processes with equidistant observations. We also see asymptotic efficiency of our estimator in the sense of minimax inequality.

We first assume following stronger conditions than [A1]–[A3] and [V].

[B1]

1. [A1] holds true with $O = \mathbb{R}^{d_2}$.
2. There exists a positive constant C such that $\sup_{t \in [0, T], \sigma \in \Lambda} |\partial_t^i \partial_x^j \partial_\sigma^k b(t, x, \sigma)| \leq C(1 + |x|)^C$ for $0 \leq 2i + j \leq 4, 0 \leq k \leq 4$ and $x \in \mathbb{R}^{d_2}$.
3. $\inf_{t, x, \sigma} \det bb^\top(t, x, \sigma) > 0$.
4. $E[|Y_0|^q] < \infty$ for any $q > 0$.
5. $\sup_t E[|\mu_t|^q] < \infty, \sup_{s < t} (E[|\mu_t - \mu_s|^q]^{1/q} (t-s)^{-1/2}) < \infty$ and $\sup_{s < t} E[(E[\mu_t - \mu_s | \mathcal{G}_s] / (t-s))^q] < \infty$.
6. For any $q > 0$, $\max_j \sup_t E[|b_t^{(j)}|^q \vee |\hat{b}_t^{(j)}|^q] < \infty$ and $\max_j \sup_{s < t} (E[|b_t^{(j)} - b_s^{(j)}|^q \vee |\hat{b}_t^{(j)} - \hat{b}_s^{(j)}|^q]^{1/q} (t-s)^{1/2}) < \infty$ for any $q > 0$.

[B2] There exist $\eta \in (0, 1/2)$, $\dot{\eta} \in (0, 1]$, $\delta > 0$ and positive-valued functions $\{a_t^j\}_{t \in [0, T], j=1,2}$ such that $b_n^{-1/2} k_n (b_n^{-1} k_n)^{\dot{\eta}} \rightarrow 0$ as $n \rightarrow \infty$, $E[\sup_{j, t > s} (|a_t^j - a_s^j|^q |t-s|^{-q\dot{\eta}})] < \infty$, $E[\sup_{j, t} |a_t^j|^q] \vee E[\sup_{j, t} (|a_t^j|^{-q})] < \infty$, and

$$\sup_n \sup_{\{[s'_{n,l}, s''_{n,l}]\} \in \mathcal{S}_\eta} E \left[\left(k_n b_n^{-1/2+\delta} \max_{1 \leq l \leq L_n} \left| b_n^{-1} (s''_{n,l} - s'_{n,l})^{-1} \# \{i; [S_{i-1}^{n,j}, S_i^{n,j}] \subset (s'_{n,l}, s''_{n,l})\} - a_{s'_{n,l}}^j \right| \right)^q \right] < \infty$$

for any $q > 0$. Moreover, there exists a positive constant γ such that $k_n b_n^{-4/7+\gamma} \rightarrow 0$ and $E[(r_n b_n^{1-\epsilon}) \vee (\underline{r}_n^{-1} b_n^{-1-\epsilon})^q] \rightarrow 0$ as $n \rightarrow \infty$ for any $q > 0$ and $\epsilon > 0$.

[B3] For any $q > 0$, there exists a positive constant c_q such that $P[\inf_{\sigma \neq \sigma_*} ((-\mathcal{Y}_0(\sigma))/|\sigma - \sigma_*|^2) \leq r^{-1}] \leq c_q/r^q$ for any $r > 0$.

[B4] There exist estimators $\{\hat{v}_n\}_{n \in \mathbb{N}}$ of v_* such that $\hat{v}_n > 0$ almost surely, $\limsup_n E[\hat{v}_n^{-q}] < \infty$, and $\sup_n E[|b_n^{1/2}(\hat{v}_n - v_*)|^q] < \infty$ for any $q > 0$.

Though Condition [B3] is rather difficult to check in a practical setting, Uchida and Yoshida [29] investigated sufficient conditions for [B3]. The simplest condition is that [B3] is satisfied if there exists $\epsilon > 0$ such that $|bb^\top(t, x, \sigma_1) - bb^\top(t, x, \sigma_2)| \geq \epsilon|\sigma_1 - \sigma_2|$ for any $t \in [0, T]$, $x \in O$ and $\sigma_1, \sigma_2 \in \Lambda$. See Remark 4 in [26] for details.

Let $U_n = \{u \in \mathbb{R}^d; \sigma_* + b_n^{-1/4}u \in \Lambda\}$, $V_n(r) = \{|u| \geq r\} \cap U_n$, and $\mathbf{Z}_n(u) = \exp(H_n(\sigma_* + b_n^{-1/4}u, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n))$ for $u \in U_n$.

Proposition 2.2 (Polynomial-type large deviation inequalities). *Assume [B1]–[B4]. Then for any $L > 0$, there exists a positive constant c_L such that $P[\sup_{u \in V_n(r)} \mathbf{Z}_n(u) \geq e^{-r/2}] \leq c_L/r^L$ for any $n \in \mathbb{N}$ and $r > 0$.*

Since $\mathbf{Z}_n(0) = 1$, Proposition 2.2 immediately yields

$$E[|b_n^{1/4}(\hat{\sigma}_n - \sigma_*)|^p] = \int_0^\infty pt^{p-1}P[|b_n^{1/4}(\hat{\sigma}_n - \sigma_*)| \geq t]dt \leq \int_0^\infty pt^{p-1}P[\sup_{u \in V_n(t)} \mathbf{Z}_n(u) \geq e^{-t/2}]dt < \infty \quad (2.10)$$

for any $p > 0$. Moreover, we obtain the following convergence of moments of the estimation error.

Theorem 2.3. *Assume [B1]–[B4]. Then $E[\mathbf{Y}f(b_n^{1/4}(\hat{\sigma}_n - \sigma_*)) \mid \mathcal{F}_1] \rightarrow E[\mathbf{Y}f(\Gamma_1^{-1/2}\mathcal{N})]$ as $n \rightarrow \infty$ for any bounded random variable \mathbf{Y} on (Ω, \mathcal{F}) and any continuous function f of at most polynomial growth.*

In particular, we obtain convergence of moments where $E[|b_n^{1/4}(\hat{\sigma}_n - \sigma_*)|^q] \rightarrow E[|\Gamma_1^{-1/2}\mathcal{N}|^q]$ for any $q > 0$. This property is used when we study the theory of information criteria and asymptotic expansion of estimators.

We also obtain results for a Bayes type estimator. Let a prior density $\pi : \Lambda \rightarrow (0, \infty)$ be a continuous function satisfying $0 < \inf_\sigma \pi(\sigma) \leq \sup_\sigma \pi(\sigma) < \infty$. Then a Bayes-type estimator $\tilde{\sigma}_n$ for the quadratic loss function is defined by

$$\tilde{\sigma}_n = \left(\int_\Lambda \exp(H_n(\sigma))\pi(\sigma)d\sigma \right)^{-1} \int_\Lambda \sigma \exp(H_n(\sigma))\pi(\sigma)d\sigma.$$

Since the Bayes-type estimator $\tilde{\sigma}_n$ contains integrals with respect to σ , we need to deal with tail behaviors of likelihood ratio $H_n(\sigma) - H_n(\sigma_*)$. Hence Proposition 2.2 is essential to deduce asymptotic properties of a Bayes-type estimator. Since the Bayes-type estimator can be calculated using Markov-Chain Monte Carlo methods, it is often easier to calculate than the maximum-likelihood-type estimator. For the Bayes-type estimator $\tilde{\sigma}_n$, we obtain similar results to the ones for the maximum-likelihood-type estimator.

Theorem 2.4. *Assume [B1]–[B4]. Then $E[\mathbf{Y}f(b_n^{1/4}(\tilde{\sigma}_n - \sigma_*)) \mid \mathcal{F}_1] \rightarrow E[\mathbf{Y}f(\Gamma_1^{-1/2}\mathcal{N})]$ as $n \rightarrow \infty$ for any bounded random variable \mathbf{Y} on (Ω, \mathcal{F}) and any continuous function f of at most polynomial growth.*

Remark 2.8. *If the assumptions of Theorem 2.2 are satisfied, asymptotic minimax theorem (Theorem 4 in [22]) holds for our model, so*

$$\lim_{\alpha \rightarrow \infty} \liminf_{n \rightarrow \infty} \sup_{|u| \leq \alpha} E_{\sigma_* + b_n^{-1/4}u} [l(|b_n^{1/4}(V_n - \sigma_* - b_n^{-1/4}u)|)] \geq E[l(|\Gamma_1\mathcal{N}|)]$$

for any estimators $\{V_n\}_n$ of the parameter and any function $l : [0, \infty) \rightarrow [0, \infty)$ which is nondecreasing and $l(0) = 0$, where E_σ denotes expectation with respect to $P_{\sigma, v_*, n}$. Using Theorems 2.3 and 2.4 and a similar argument in Theorem 2.2 of Ogihara [25], we can see that $\hat{\sigma}_n$ and $\tilde{\sigma}_n$ attain the lower bound of the above inequality for continuous l of at most polynomial growth, if further [B2] and uniform versions of [B3] and [B4] with respect to the true value (σ_*, v_*) are satisfied. Hence our estimators are asymptotically efficient in this sense as well.

3 Simulation results

In this section, we examine some simulation results of our estimator.

First, we consider the case where the latent process Y is a Brownian motion, that is, Y satisfies the following stochastic differential equation:

$$\begin{cases} dY_t^1 &= \sigma_{1,*}dW_t^1 \\ dY_t^2 &= \sigma_{3,*}dW_t^1 + \sigma_{2,*}dW_t^2, \end{cases}$$

Table 1: Simulation results for estimators of parameters

n		Results with $v_* = (0.001, 0.001)$					Results with $v_* = (0.005, 0.005)$				
		σ_1	σ_2	σ_3	v_1	v_2	σ_1	σ_2	σ_3	v_1	v_2
1000	$(\hat{\sigma}_n, \hat{v}_n)$	0.897 (0.040)	0.776 (0.042)	0.451 (0.062)	0.001504 (0.000079)	0.001500 (0.000080)	0.957 (0.086)	0.818 (0.143)	0.481 (0.094)	0.005515 (0.000293)	0.005501 (0.000296)
	$(\hat{\sigma}'_n, \hat{v}'_n)$	0.971 (0.046)	0.840 (0.047)	0.487 (0.067)	0.001100 (0.000075)	0.001094 (0.000078)	0.991 (0.092)	0.850 (0.139)	0.498 (0.098)	0.005053 (0.000298)	0.005035 (0.000306)
	$\hat{\sigma}''_n$	0.999 (0.045)	0.863 (0.046)	0.501 (0.068)	-	-	0.997 (0.069)	0.861 (0.070)	0.499 (0.096)	-	-
	$(\hat{\sigma}_n, \hat{v}_n)$	0.964 (0.028)	0.833 (0.029)	0.481 (0.040)	0.001099 (0.000026)	0.001099 (0.000026)	0.990 (0.044)	0.854 (0.044)	0.495 (0.061)	0.005095 (0.000121)	0.005096 (0.000123)
	$(\hat{\sigma}'_n, \hat{v}'_n)$	0.997 (0.031)	0.862 (0.031)	0.498 (0.041)	0.001006 (0.000027)	0.001006 (0.000027)	0.999 (0.045)	0.862 (0.045)	0.499 (0.062)	0.004996 (0.000123)	0.004998 (0.000125)
	$\hat{\sigma}''_n$	0.999 (0.029)	0.864 (0.030)	0.499 (0.041)	-	-	0.998 (0.043)	0.862 (0.044)	0.499 (0.062)	-	-
true values		1	0.866	0.5	0.001	0.001	1	0.866	0.5	0.005	0.005

where $\sigma_* = (\sigma_{1,*}, \sigma_{2,*}, \sigma_{3,*}) \in (\epsilon, R) \times (-R, R) \times (\epsilon, R)$ for some $0 < \epsilon < R$. Moreover, let $\{N_t^1\}_{0 \leq t \leq T}$ and $\{N_t^2\}_{0 \leq t \leq T}$ be two independent Poisson processes with parameters λ_1 and λ_2 , respectively. We give sampling times by $S_i^{n,j} = \inf\{N_{nt}^j \geq j\} \wedge T$ for $j = 1, 2$. Let $\{\epsilon_i^{n,j}\}_{i \in \mathbb{Z}_+, j=1,2}$ be independent normal random variables with $E[\epsilon_i^{n,j}] = 0$ and $E[(\epsilon_i^{n,j})^2] = v_{j,*}$.

Then we can see that this example satisfies [A1''], [A2] and [A3']. So the maximum-likelihood-type estimator $\hat{\sigma}_n$ is asymptotically mixed normal and asymptotically efficient with asymptotic variance Γ_1^{-1} . For the estimator \hat{v}_n of v_* we first use a simple estimator $\hat{v}_n = (2\mathbf{J}_{k,n})^{-1} \sum_i (\tilde{Y}_i^k - \tilde{Y}_{i-1}^k)^2$, which means that our estimator is calculated by $\hat{\sigma}_n = \arg\max_{\sigma} H_n(\sigma, \hat{v}_n)$. We also consider a plug-in estimator $\hat{v}'_{k,n} = (\hat{v}_{k,n} - |b^k(\hat{\sigma}_n)|^2 T / (2\mathbf{J}_{k,n})) \vee 0$ of $v_{k,*}$, and $\hat{\sigma}'_n = \arg\max_{\sigma} H_n(\sigma, \hat{v}'_n)$. Let $\hat{\sigma}''_n = \arg\max_{\sigma} H_n(\sigma, v_*)$. Then $\hat{\sigma}''_n$ cannot be calculated by observed data, but we can use it for comparison. Though these estimators have the same asymptotic variance, their performances for finite samples are different. In particular, we cannot ignore the bias of \hat{v}_n since v is relatively small compared with σ in practical data.

Table 1 shows results of 1000 estimations. Each cell represents the average of estimators, with sample standard deviations given in parentheses. We set the values of parameters as $k_n = \lceil n^{5/8} \rceil$, $T = 1$, $(\lambda_1, \lambda_2) = (1, 1)$, $(\sigma_{1,*}, \sigma_{2,*}, \sigma_{3,*}) = (1, \sqrt{1 - 0.5^2}, 0.5)$, and consider two cases of the noise variances: $v_* = (0.001, 0.001)$ for the left-hand side of the table and $v_* = (0.005, 0.005)$ for the right-hand side. In both cases, we can see that \hat{v}_n has an upper bias for $n = 1000$, and causes a lower bias of $\hat{\sigma}_n$ because \hat{v}_n contains variance of the latent process, which is always positive. These biases can be moderated by using the plug-in estimator. For $n = 5000$, the plug-in estimator $\hat{\sigma}'_n$ performs as well as $\hat{\sigma}''_n$. In the case of $v_* = (0.005, 0.005)$, the biases of \hat{v}_n and \hat{v}'_n are relatively small, so the performance of $\hat{\sigma}_n$ and $\hat{\sigma}'_n$ are better.

We can also construct an estimator $\hat{\sigma}'_{1,n} \hat{\sigma}'_{3,n} T$ of the quadratic covariation $\langle Y^1, Y^2 \rangle_T = \sigma_{1,*} \sigma_{3,*} T$. We see that

$$n^{1/4}(\hat{\sigma}'_{1,n} \hat{\sigma}'_{3,n} T - \langle Y^1, Y^2 \rangle_T) \rightarrow^d N(0, \sigma_{3,*}^2 (\Gamma_1^{-1})_{11} + 2\sigma_{1,*} \sigma_{3,*} (\Gamma_1^{-1})_{13} + \sigma_{1,*}^2 (\Gamma_1^{-1})_{33}) \quad (3.1)$$

as $n \rightarrow \infty$ by the delta method, and the estimator is asymptotically efficient since we can reparameterize the model using $\sigma_{1,*} \sigma_{3,*}$. We therefore compared the performance of the estimator (MLE) with existing estimators of the quadratic covariation. We used the pre-averaged Hayashi–Yoshida estimator (PHY) and modulated realized covariance (MRC) by Christensen, Kinnebrock, and Podolskij [7], the local method of moments (LMM) by Bibinger et al. [5], and an estimator based on maximum likelihood estimator of a model of constant diffusion coefficients (QMLE) by Ait-Sahalia, Fan, and Xiu [2] for comparison. Except LMM these estimators can be calculated using the ‘cce’ function in the ‘yuima’ R package (<http://r-forge.r-project.org/projects/yuima>). We used the default values of the ‘cce’ function or values used in corresponding papers for parameters of estimators ($\theta = 0.15$ for PHY, $\theta = 1$ for MRC₁, $J = 30$, $h^{-1} = 10$ for LMM). Here we use the *oracle estimator* defined in [5] for LMM to avoid a complicated calculation. For the modulated realized covariance, we also examine an estimator MRC₂ with $\theta = 1/3$ which is used in Jacod et al. [20]. Table 2 shows the results of 1000 estimations. We used the same parameter values as above. Then the true value of the quadratic covariation becomes $\langle Y^1, Y^2 \rangle_T = 0.5$. For both cases of observation noise variance, we can see that sample standard deviations of

Table 2: Comparison of estimators of $\langle Y^1, Y^2 \rangle_T$

n	Results with $v_* = (0.001, 0.001)$						Theoretical minimum
	MLE	PHY	MRC ₁	MRC ₂	QMLE	LMM	
1000	0.474 (0.073)	0.499 (0.121)	0.508 (0.182)	0.501 (0.110)	0.501 (0.095)	0.463 (0.082)	(0.066)
5000	0.496 (0.046)	0.497 (0.081)	0.504 (0.124)	0.499 (0.073)	0.498 (0.056)	0.497 (0.069)	(0.044)

n	Results with $v_* = (0.005, 0.005)$						Theoretical minimum
	MLE	PHY	MRC ₁	MRC ₂	QMLE	LMM	
1000	0.496 (0.109)	0.497 (0.148)	0.508 (0.185)	0.5000 (0.124)	0.5000 (0.120)	0.518 (0.112)	(0.099)
5000	0.499 (0.069)	0.497 (0.098)	0.505 (0.126)	0.499 (0.083)	0.499 (0.079)	0.514 (0.083)	(0.066)

Table 3: Estimation errors of estimators of $\langle Y^1, Y^2 \rangle_T$ for the CIR process

n	MLE	PHY	MRC ₁	MRC ₂	QMLE	LMM
1000	-0.0267 (0.0733)	-0.0063 (0.1286)	-0.0058 (0.1867)	-0.0036 (0.1162)	-0.0008 (0.1013)	-0.0348 (0.0844)
5000	-0.0023 (0.0456)	-0.0036 (0.0858)	-0.0022 (0.1305)	-0.0016 (0.0768)	-0.0005 (0.0580)	-0.0033 (0.0719)

our estimator are the best in large samples. The theoretical (asymptotic) minimum of standard deviations for all estimators is calculated as $n^{-1/4}(\sigma_{3,*}^2(\Gamma_1^{-1})_{11} + 2\sigma_{1,*}\sigma_{3,*}(\Gamma_1^{-1})_{13} + \sigma_{1,*}^2(\Gamma_1^{-1})_{33})^{1/2}$. Table 2 also shows that the sample standard deviations of MLE are close to the minima in large samples.

In the next, we consider the model with random diffusion coefficients and non-Gaussian noise. As mentioned in Remark 2.1, we cannot directly apply our results to stochastic volatility models. Here we consider the Cox-Ingersoll-Ross (CIR) process derived in [9] as a latent process with random diffusion coefficients. Let the latent process Y satisfy

$$dY_t = \begin{pmatrix} \alpha_1 - \beta_1 Y_t^1 \\ \alpha_2 - \beta_2 Y_t^2 \end{pmatrix} dt + \begin{pmatrix} \sigma_{1,*}\sqrt{Y_t^1} & 0 \\ \sigma_{3,*}\sqrt{Y_t^2} & \sigma_{2,*}\sqrt{Y_t^2} \end{pmatrix} dW_t,$$

where $\sigma_* = (\sigma_{1,*}, \sigma_{2,*}, \sigma_{3,*}) \in (\epsilon', R') \times (-R', R') \times (\epsilon', R')$. We assume Conditions $2\alpha_1 > \sigma_{1,*}^2$ and $2\alpha_2 > \sigma_{2,*}^2 + \sigma_{3,*}^2$ which ensure $Y_t^1 > 0$ and $Y_t^2 > 0$ for $t \in [0, T]$ almost surely. Let $\{\epsilon_i^{n,j}\}_{i \in \mathbb{Z}}$ be i.i.d. random variables following a centered Gamma distribution with a shape parameter k_j and a scale parameter θ_j for $j = 1, 2$. We define $\{N_t^j\}$, \hat{v}_n , \hat{v}'_n , $\hat{\sigma}_n$, and $\hat{\sigma}'_n$ similarly to the first example. We set the values of parameters as $k_n = \lceil n^{5/8} \rceil$, $T = 1$, $(\lambda_1, \lambda_2) = (1, 1)$, $(\sigma_{1,*}, \sigma_{2,*}, \sigma_{3,*}) = (1, \sqrt{1 - 0.5^2}, 0.5)$, $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (1, 1, 1, 1)$, and $(k_1, k_2, \theta_1, \theta_2) = (2, 2, \sqrt{0.0005}, \sqrt{0.0005})$ which implies $v_* = (0.001, 0.001)$. Table 3 shows averages and sample standard deviations of $T_n - \langle Y^1, Y^2 \rangle_T$ for each estimator T_n of the quadratic covariation $\langle Y^1, Y^2 \rangle_T$ in 1000 simulations. $\langle Y^1, Y^2 \rangle_T$ is random in this model since the diffusion coefficients are random. So we use extra-high-frequency observations $\{Y_{k/100000}^l\}_{k=0}^{100000}$ of Y to calculate the approximated true value of $\langle Y^1, Y^2 \rangle_T$. In this model, we have not obtained the LAMN property nor asymptotic efficiency of our estimator though we expect to obtain them. However, we still see that our estimator achieves the best error variance in large samples.

4 Asymptotically equivalent representation of the quasi-likelihood function

We will prove our main results in the rest of this paper. In this section, we introduce an asymptotically equivalent representation $\tilde{H}_n(\sigma, v)$ of the quasi-likelihood function $H_n(\sigma, v)$, and prove the equivalence. \tilde{H}_n is a

useful function for deducing the limit of H_n .

4.1 Some notations

We denote E_m as the $\mathcal{G}_{s_{m-1}}$ -conditional expectation and $\bar{E}_m[\mathbf{X}] = \mathbf{X} - E_m[\mathbf{X}]$ for a random variable \mathbf{X} . We use the symbol C for a generic positive constant that can vary from line to line.

For a sequence c_n of positive-valued $\mathfrak{B}(\Pi_n)$ -measurable random variables, let us denote by $\{\bar{R}_n(c_n)\}_{n \in \mathbb{N}}$, $\{\underline{R}_n(c_n)\}_{n \in \mathbb{N}}$ and $\{\dot{R}_n(c_n)\}_{n \in \mathbb{N}}$ sequences of random variables (which may depend on $1 \leq m \leq \ell_n$ and σ) satisfying

$$E[(c_n^{-1}(r_n/b_n)^{-p_1}(b_n/\underline{r}_n)^{-p_2}(\bar{k}_n/k_n)^{-p_3}(k_n/\underline{k}_n)^{-p_4}b_n^{-\delta} \sup_{\sigma, m} E_{\Pi} [|\bar{R}_n(c_n)|^q]^{1/q})^{q'}] \rightarrow 0,$$

$$E[(c_n^{-1}(r_n/b_n)^{q_1}(b_n/\underline{r}_n)^{q_2}(\bar{k}_n/k_n)^{q_3}(k_n/\underline{k}_n)^{q_4}b_n^{\delta'} \sup_{\sigma, m} E_{\Pi} [|\underline{R}_n(c_n)|^q]^{1/q})^{q'}] \rightarrow 0,$$

and

$$c_n^{-1}(r_n/b_n)^{q_1}(b_n/\underline{r}_n)^{q_2}(\bar{k}_n/k_n)^{q_3}(k_n/\underline{k}_n)^{q_4} \sup_{\sigma, m} |\dot{R}_n(c_n)| \rightarrow^p 0,$$

respectively, as $n \rightarrow \infty$ for any $\delta, q, q', q_1, \dots, q_4 > 0$ with some constants $\delta', p_1, \dots, p_4 \geq 0$.

Let $M_m(v) = \text{diag}(v_1 M_{1,m}, v_2 M_{2,m})$ for $v = (v_1, v_2)$, $\tilde{b}_m^k = b^k(s_{m-1}, X_{s_{m-1}}, \sigma)$, $\tilde{b}_{m,*}^k = b^k(s_{m-1}, X_{s_{m-1}}, \sigma_*)$,

$$\tilde{Z}_m = (((\tilde{b}_{m,*}^1 \cdot (W_{S_i^{n,1}} - W_{S_{i-1}^{n,1}}) + \epsilon_i^{n,1} - \epsilon_{i-1}^{n,1})_{i=K_{m-1}^1+2}^{K_m^1})^\top, ((\tilde{b}_{m,*}^2 \cdot (W_{S_j^{n,2}} - W_{S_{j-1}^{n,2}}) + \epsilon_j^{n,2} - \epsilon_{j-1}^{n,2})_{j=K_{m-1}^2+2}^{K_m^2})^\top)^\top,$$

$$\tilde{S}_m(\sigma, v) = \begin{pmatrix} \text{diag}(|\tilde{b}_m^1|^2 |I_{i,m}^1|)_i & \{\tilde{b}_m^1 \cdot \tilde{b}_m^2 |I_{i,m}^1 \cap I_{j,m}^2|_{ij}\} \\ \{\tilde{b}_m^1 \cdot \tilde{b}_m^2 |I_{i,m}^1 \cap I_{j,m}^2|_{ji}\} & \text{diag}(|\tilde{b}_m^2|^2 |I_{j,m}^2|)_j \end{pmatrix} + M_m(v), \quad (4.1)$$

and

$$\tilde{H}_n(\sigma, v) = -\frac{1}{2} \sum_{m=2}^{\ell_n} \tilde{Z}_m^\top \tilde{S}_m^{-1}(\sigma, v) \tilde{Z}_m - \frac{1}{2} \sum_{m=2}^{\ell_n} \log \det \tilde{S}_m(\sigma, v).$$

The diffusion coefficients b in \tilde{Z}_m and \tilde{S}_m are either $b(s_{m-1}, X_{s_{m-1}}, \sigma)$ or $b(s_{m-1}, X_{s_{m-1}}, \sigma_*)$. Hence we do not need to consider the time-dependent structure of b when we study asymptotics of the summands in \tilde{H}_n . In particular, we obtain $E_m[\tilde{Z}_m^\top \partial_\sigma \tilde{S}_m(\sigma_*, v_*)^{-1} \tilde{Z}_m + \partial_\sigma \log \det \tilde{S}_m(\sigma_*, v_*)] = 0$ by $\partial_\sigma \log \det \tilde{S}_m(\sigma, v) = -\text{tr}(\partial_\sigma \tilde{S}_m \tilde{S}_m^{-1})(\sigma, v)$. We will prove the asymptotic equivalence of H_n and \tilde{H}_n and then investigate asymptotic properties of H_n instead of \tilde{H}_n .

Similarly to the approach of Gloter and Jacod [13], we first show our results under the following condition [A1'], which is stronger than [A1]. Then localization techniques and Girsanov's theorem enable us to replace [A1'] with [A1].

[A1'] Condition [A1] is satisfied, $O = \mathbb{R}^{d_2}$, $\sup_{t,x,\sigma} \|(bb^\top)^{-1}\|(t, x, \sigma) < \infty$, $\mu_t \equiv 0$ and $Y_0, \sup_t |b_t^{(l)}|, \partial_t^i \partial_x^j \partial_\sigma^k b$, and $\sup_{t>s} ((|b_t^{(l)} - b_s^{(l)}| \vee |\hat{b}_t^{(l)} - \hat{b}_s^{(l)}|)/(t-s))$ are all bounded for $l = 0, 1$, $0 \leq 2i + j \leq 4$ and $0 \leq k \leq 4$.

We can also see that [A1'] implies [B1].

4.2 Fundamental properties of the noise covariance matrix

In the following subsection we will show the asymptotic equivalence of \tilde{H}_n and H_n , namely that

$$b_n^{-1/2} \sup_{\sigma \in \Lambda} |\partial_\sigma^j (H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n)) - \partial_\sigma^j (\tilde{H}_n(\sigma, v_*) - \tilde{H}_n(\sigma_*, v_*))| \rightarrow^p 0$$

as $n \rightarrow \infty$ for $0 \leq j \leq 3$. To that end, we first show fundamental properties of S_m and \tilde{S}_m . These matrices inherit some properties of $M_{j,m}$, that are necessary to deduce the limit of H_n and \tilde{H}_n . The first property (4.2) concerns the trace of a matrix related to $M_{j,m}$ investigated by [13]. In the one-dimensional model with noisy, equidistance observations, this property can be directly applied to the quasi-likelihood function because the covariance matrix of the latent process is the unit matrix. However, this is insufficient for our purpose because

our covariance matrix $S_m - M_m(v)$ of the latent process is rather complicated. Therefore, we investigate further matrix properties related to $M_{j,m}$.

First, we consider the results in [13]. For any positive constants p, q, a and b , eigenvalues of $(a\mathcal{E} + M_{j,m})^{-1}$ are $\{(a + 2(1 - \cos(i\pi(k_m^j + 1)^{-1}))\}_{i=1}^{k_m^j}$ and we obtain

$$\pi^{-1}k_m^j I_p(a) - a^{-p} \leq \text{tr}((a\mathcal{E} + M_{j,m})^{-p}) \leq \pi^{-1}k_m^j I_p(a), \quad (4.2)$$

$$\pi^{-1}k_m^j I_{p,q}(a, b) - a^{-p}b^{-q} \leq \text{tr}((a\mathcal{E} + M_{j,m})^{-p}(b\mathcal{E} + M_{j,m})^{-q}) \leq \pi^{-1}k_m^j I_{p,q}(a, b), \quad (4.3)$$

where $I_p(a) = \int_0^\pi (a + 2(1 - \cos x))^{-p} dx$ and $I_{p,q}(a, b) = \int_0^\pi (a + 2(1 - \cos x))^{-p} (b + 2(1 - \cos x))^{-q} dx$. Simple calculations show that $I_1(a) = \pi/\sqrt{a(4+a)}$, $I_2(a) = \pi(2+a)a^{-3/2}(4+a)^{-3/2}$ and $\int_0^\pi \{\log(a + 2(1 - \cos x)) - \log(b + 2(1 - \cos x))\} dx = 2\pi(\log(\sqrt{a} + \sqrt{4+a}) - \log(\sqrt{b} + \sqrt{4+b}))$. See Section 4.1 in [13] for the details. Moreover, differentiation with respect to a yields

$$I_p(a) = \frac{(-1)^{p-1}}{(p-1)!} \left(\frac{d}{da} \right)^{p-1} \left(\frac{\pi}{\sqrt{a(4+a)}} \right).$$

In particular, if $a = \mathbf{X}_n b_n^{-1}$ for some tight random variables $\{\mathbf{X}_n\}_n$, then we have

$$I_p(\mathbf{X}_n b_n^{-1}) = \frac{\pi(2p-3)!!}{2^p(p-1)!} (\mathbf{X}_n b_n^{-1})^{-p+1/2} + O_p(b_n^{p-3/2}).$$

For $\epsilon \geq 0$, let $\{p_j(\epsilon)\}_{j \in \mathbb{N}}$ and $\{p'_j(\epsilon)\}_{j \in \mathbb{N}}$ be sequences of positive numbers satisfying $p_1(\epsilon) = 2 + \epsilon$, $p'_1(\epsilon) = 1 + \epsilon$, $p_{j+1}(\epsilon) = 2 + \epsilon - 1/p_j(\epsilon)$, and $p'_{j+1}(\epsilon) = 2 + \epsilon - 1/p'_j(\epsilon)$ for $j \in \mathbb{N}$. Let $E_{i,j}(a)$ be a $k_m^j \times k_m^j$ matrix satisfying $(E_{i,j}(a))_{k,l} = \delta_{k,l} + a\delta_{(i,j)}(k, l)$ for $a \in \mathbb{R}$. Then we have

$$E_{k_m^j, k_m^j - 1}(p_{k_m^j - 1}(\epsilon)^{-1}) \cdots E_{2,1}(p_1(\epsilon)^{-1})(\epsilon\mathcal{E} + M_{j,m})E_{1,2}(p_1(\epsilon)^{-1}) \cdots E_{k_m^j - 1, k_m^j}(p_{k_m^j - 1}(\epsilon)^{-1}) = \text{diag}((p_j(\epsilon))_{j=1}^{k_m^j}),$$

$$(\epsilon\mathcal{E} + M_{j,m})^{-1} = E_{1,2}(p_1(\epsilon)^{-1}) \cdots E_{k_m^j - 1, k_m^j}(p_{k_m^j - 1}(\epsilon)^{-1}) \text{diag}((p_j(\epsilon)^{-1})_{j=1}^{k_m^j}) E_{k_m^j, k_m^j - 1}(p_{k_m^j - 1}(\epsilon)^{-1}) \cdots E_{2,1}(p_1(\epsilon)^{-1}),$$

and hence

$$(\epsilon\mathcal{E} + M_{j,m})^{-1} = \left\{ \prod_{k+1 \leq i \leq l} p_{i-1}(\epsilon)^{-1} 1_{\{k \leq l\}} \right\}_{k,l} \text{diag}((p_j(\epsilon)^{-1})_{j=1}^{k_m^j}) \left\{ \prod_{l+1 \leq i \leq k} p_{i-1}(\epsilon)^{-1} 1_{\{l \leq k\}} \right\}_{k,l}. \quad (4.4)$$

Moreover, we have the following lemma.

Lemma 4.1. *Let $\epsilon \in [0, 1)$ and $p_+(\epsilon) = 1 + \epsilon/2 + \sqrt{\epsilon + \epsilon^2/4}$. Then*

1. $1 \leq p'_j(\epsilon) \leq p_+(\epsilon) < p_j(\epsilon) \leq 1 + 1/j + j\epsilon$ for $j \in \mathbb{N}$, $\{p_j(\epsilon)\}_j$ is monotone decreasing, and $\{p'_j(\epsilon)\}_j$ is monotone nondecreasing.
2. $\{((\epsilon\mathcal{E} + M_{j,m})^{-1})_{kk}\}_{k=1}^{[k_m^j/2]}$ is monotone increasing.
3. $p_j - p_+ \leq (1 + \sqrt{\epsilon})^{-(j-2)}$ and $p_+ - p'_j \leq \sqrt{\epsilon}(1 + \sqrt{\epsilon})^{-(j-2)}$ for $j \geq 2$.
4. $\prod_{j=1}^k p'_j(\epsilon) = (p_k(\epsilon) - 1) \prod_{j=1}^{k-1} p_j(\epsilon)$ for any $k \geq 2$.

Proof. 1. We simply denote $p_j = p_j(\epsilon)$. We will prove $p_+(\epsilon) < p_j(\epsilon) \leq 1 + 1/j + j\epsilon$ for $j \in \mathbb{N}$ by induction. The results obviously hold for $j = 1$. Assume the results hold for all values in \mathbb{N} up to j . Then since $p_+ = 2 + \epsilon - 1/p_+$, we obtain $p_{j+1} - p_+ = 1/p_+ - 1/p_j > 0$, and

$$p_{j+1} \leq 2 + \epsilon - (j/(j+1))(1 + j^2\epsilon/(j+1))^{-1} \leq 2 + \epsilon - (j/(j+1))(1 - j^2\epsilon/(j+1)) \leq 1 + 1/(j+1) + (j+1)\epsilon.$$

Hence, we have $p_+(\epsilon) < p_j(\epsilon) \leq 1 + 1/j + j\epsilon$ for $j \in \mathbb{N}$. Moreover, we can inductively deduce $p_{j+1} - p_j = 1/p_{j-1} - 1/p_j > 0$. The results for $\{p'_j(\epsilon)\}_j$ are obtained similarly.

2. By considering the cofactor matrix and (4.4), we have

$$((\epsilon\mathcal{E} + M_{j,m})^{-1})_{kk} = \frac{\det(\epsilon\mathcal{E}_{k-1} + M(k-1)) \det(\epsilon\mathcal{E}_{k_m^j-k} + M(k_m^j - k))}{\det(\epsilon\mathcal{E} + M_{j,m})} = \frac{\prod_{l=1}^{k-1} p_l \prod_{l=1}^{k_m^j-k} p_l}{\prod_{l=1}^{k_m^j} p_l}. \quad (4.5)$$

Therefore we obtain the result by monotonicity of p_j .

3. This is easy since $p_j - p_+ = (p_{j-1} - p_+)/p_+ p_{j-1} \leq (p_1 - p_+)/p_+^{j-1} \leq p_+^{-j+2}$.

4.

$$\prod_{j=1}^k p'_j(\epsilon) = \det(\epsilon\mathcal{E} + M(k) - (E_{11}(1) - \mathcal{E})) = \det(\epsilon\mathcal{E} + M(k) - (E_{kk}(1) - \mathcal{E})) = (p_k(\epsilon) - 1) \prod_{j=1}^{k-1} p_j(\epsilon).$$

□

Let $\bar{\rho} = \sup_{t,\sigma} (|b^1 \cdot b^2| |b^1|^{-1} |b^2|^{-1})(t, X_t, \sigma)$, $\tilde{D}_m = (\tilde{D}_{1,m}, \tilde{D}_{2,m})$, $\tilde{D}_{j,m} = \text{diag}(|\tilde{b}_m^j|^2 |I_{i,m}^j|) + v_{j,*} M_{j,m}$, $D'_m = (D'_{1,m}, D'_{2,m})$, $D'_{j,m} = \text{diag}(|I_{i,m}^j|)$, and $\tilde{D}_{j,m} = |\tilde{b}_m^j|^2 r_n \mathcal{E} + v_{j,*} M_{j,m}$.

Lemma 4.2. Assume [B1]. Then $\text{tr}(\tilde{S}_m^{-1}(\sigma, v_*)) = \bar{R}_n(b_n^{1/2} k_n)$.

Proof. Let $D''_m = \text{diag}(|\tilde{b}_m^1|^2 D'_{1,m}, |\tilde{b}_m^2|^2 D'_{2,m})$ and $D'''_m = (D''_m)^{-1/2} \tilde{D}_m (D''_m)^{-1/2}$, then we have

$$\tilde{S}_m = (D''_m)^{1/2} (D'''_m)^{1/2} (\mathcal{E} + (D''_m)^{-1/2} (D'''_m)^{-1/2} (\tilde{S}_m - \tilde{D}_m) (D''_m)^{-1/2} (D'''_m)^{-1/2}) (D'''_m)^{1/2} (D''_m)^{1/2}.$$

Moreover, Lemma A.4, [B1], and Lemma 2 in [26] yield

$$\begin{aligned} & \left\| (D'''_m)^{-1/2} (D''_m)^{-1/2} (\tilde{S}_m - \tilde{D}_m) (D''_m)^{-1/2} (D'''_m)^{-1/2} \right\| \\ & \leq \left\| (D''_m)^{-1/2} (\tilde{S}_m - \tilde{D}_m) (D''_m)^{-1/2} \right\| \leq \bar{\rho} \left\{ \left\| \left\{ \frac{|I_{i,m}^1 \cap I_{j,m}^2|}{|I_{i,m}^1|^{1/2} |I_{j,m}^2|^{1/2}} \right\}_{i,j} \right\| \vee \left\| \left\{ \frac{|I_{i,m}^1 \cap I_{j,m}^2|}{|I_{i,m}^1|^{1/2} |I_{j,m}^2|^{1/2}} \right\}_{j,i} \right\| \right\} \leq \bar{\rho} < 1. \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} \text{tr}(\tilde{S}_m^{-1}) & \leq \text{tr}((D'''_m)^{-1/2} (D''_m)^{-1} (D'''_m)^{-1/2}) \|(\mathcal{E} + (D''_m)^{-1/2} (D'''_m)^{-1/2} (\tilde{S}_m - \tilde{D}_m) (D''_m)^{-1/2} (D'''_m)^{-1/2})^{-1}\| \\ & \leq \text{tr}(\tilde{D}_m^{-1}) / (1 - \bar{\rho}) \leq \sum_{j=1}^2 \text{tr}(\tilde{D}_{j,m}^{-1}) r_n \underline{\mathbf{L}}^{-1} (1 - \bar{\rho})^{-1}, \end{aligned}$$

by Lemma A.1, the equation $\tilde{D}_{j,m} = \tilde{D}_{j,m}^{1/2} (\mathcal{E} - \tilde{D}_{j,m}^{-1/2} (\tilde{D}_{j,m} - \tilde{D}_{j,m}) \tilde{D}_{j,m}^{-1/2}) \tilde{D}_{j,m}^{1/2}$, and that

$$\|\tilde{D}_{j,m}^{-1/2} (\tilde{D}_{j,m} - \tilde{D}_{j,m}) \tilde{D}_{j,m}^{-1/2}\| \leq (|\tilde{b}_m^j|^2 r_n)^{-1} |\tilde{b}_m^j|^2 (r_n - \underline{\mathbf{L}}) = 1 - \underline{\mathbf{L}}/r_n. \quad (4.6)$$

We thus obtain the results by (4.2). □

4.3 Asymptotic equivalence of H_n and \tilde{H}_n

In this section, we prove the asymptotic equivalence of H_n and \tilde{H}_n . We provide the following lemma about estimates of moments of the quantities related to H_n and \tilde{H}_n . The proof is given in the appendix; it is obtained based on the properties of $M_{j,m}$ in Section 4.2, standard Itô calculus, and some results from linear algebra.

Let $\tilde{S}_{m,*} = \tilde{S}_m(\sigma_*, v_*)$ and

$$\mathbf{S}(t, x, \sigma, v) = \begin{pmatrix} \{|b^1(t, x, \sigma)|^2 |I_{i,m}^1| \delta_{ii'}\}_{ii'} + v_1 M_{1,m} & \{b^1 \cdot b^2(t, x, \sigma) |I_{i,m}^1 \cap I_{j,m}^2\}_{ij} \\ \{b^1 \cdot b^2(t, x, \sigma) |I_{i,m}^1 \cap I_{j,m}^2\}_{ji} & \{|b^2(t, x, \sigma)|^2 |I_{j,m}^2| \delta_{jj'}\}_{jj'} + v_2 M_{2,m} \end{pmatrix}.$$

Lemma 4.3. Assume [B1]. Let $\sigma \in \Lambda$, $k_1, k_2, k_3 \in \mathbb{Z}_+$, $k_1 + k_2 \geq 1$, $k_1 \leq 4$, $k_2 \leq 4$, \mathbf{X}_m be a $\mathcal{G}_{s_{m-1}}$ -measurable random variable, and $\mathbf{S}' = \partial_\sigma^{k_1} \partial_x^{k_2} \partial_v^{k_3} \mathbf{S}^{-1}(s_{m-1}, \mathbf{X}_m, \sigma, v_*)$. Then

1. $E_m[(\tilde{Z}_m^\top \mathbf{S}' \tilde{Z}_m)^2] = 2\text{tr}((\mathbf{S}' \tilde{S}_{m,*})^2) + \text{tr}(\mathbf{S}' \tilde{S}_{m,*})^2 + \bar{R}_n(1)$, $E_m[(\tilde{Z}_m^\top \mathbf{S}' \tilde{Z}_m)^4] = \bar{R}_n((b_n^{-4} k_n^7) \vee (b_n^{-2} k_n^4))$ and $E_m[(\tilde{Z}_m^\top \mathbf{S}' \tilde{Z}_m)^q] = \bar{R}_n(b_n^{-q} k_n^{2q})$ for $q > 4$.
2. $E_\Pi[(\sum_m (Z_m - \tilde{Z}_m)^\top \mathbf{S}' (Z_m + \tilde{Z}_m))^q] = \bar{R}_n((b_n^{-3} k_n^7)^{q/4})$ for $q \geq 4$.
3. $E_\Pi[(\sum_m (Z_m - \tilde{Z}_m)^\top \mathbf{S}' (Z_m + \tilde{Z}_m))^2] = \bar{R}_n((b_n^{-1} k_n^2) \vee (b_n^{-2} k_n^{7/2}))$.

Proof. See the appendix. \square

Now we obtain the asymptotic equivalence of H_n and \tilde{H}_n .

Lemma 4.4. Assume [B1], [A2], and [V]. Then

$$b_n^{-1/2} \sup_{\sigma \in \Lambda} |\partial_\sigma^j (H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n)) - \partial_\sigma^j (\tilde{H}_n(\sigma, v_*) - \tilde{H}_n(\sigma_*, v_*))| \rightarrow^p 0,$$

and $b_n^{-1/4} (\partial_\sigma H_n(\sigma_*, \hat{v}_n) - \partial_\sigma \tilde{H}_n(\sigma_*, v_*)) \rightarrow^p 0$ as $n \rightarrow \infty$ for $0 \leq j \leq 3$. If [B4] holds as well, then

$$E_\Pi \left[\left(b_n^{-1/2} \sup_{\sigma \in \Lambda} |\partial_\sigma^j (H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n)) - \partial_\sigma^j (\tilde{H}_n(\sigma, v_*) - \tilde{H}_n(\sigma_*, v_*))| \right)^q \right] = \bar{R}_n((b_n^{-5} k_n^7)^{q/4})$$

for any $0 \leq j \leq 3$ and $q > 0$.

Proof. We first obtain

$$\begin{aligned} & H_n(\sigma, v_*) - H_n(\sigma_*, v_*) - (\tilde{H}_n(\sigma, v_*) - \tilde{H}_n(\sigma_*, v_*)) \\ &= -\frac{\sigma - \sigma_*}{2} \sum_m \left\{ (Z_m - \tilde{Z}_m)^\top \int_0^1 \partial_\sigma S_m^{-1}(\sigma_t, v_*) dt (Z_m + \tilde{Z}_m) + \tilde{Z}_m \int_0^1 (\partial_\sigma S_m^{-1}(\sigma_t, v_*) - \partial_\sigma \tilde{S}_m^{-1}(\sigma_t, v_*)) dt \tilde{Z}_m \right. \\ & \quad \left. + \int_0^1 \partial_\sigma \log \frac{\det S_m(\sigma_t, v_*)}{\det \tilde{S}_m(\sigma_t, v_*)} dt \right\} =: \hat{\Psi}_{1,n}(\sigma) + \hat{\Psi}_{2,n}(\sigma) + \hat{\Psi}_{3,n}(\sigma). \end{aligned}$$

We will give estimates for these quantities. Point 2 of Lemma 4.3 yields $\sup_\sigma E_\Pi[|b_n^{-1/2} \partial_\sigma^j \hat{\Psi}_{1,n}|^q] = \bar{R}_n((b_n^{-5} k_n^7)^{q/4})$ for $0 \leq j \leq 4$ and $q > 0$, and consequently by Sobolev's inequality $E_\Pi[\sup_\sigma |b_n^{-1/2} \partial_\sigma^j \hat{\Psi}_{1,n}|^q] = \bar{R}_n((b_n^{-5} k_n^7)^{q/4})$ as $n \rightarrow \infty$ for $0 \leq j \leq 3$ and $q > 0$.

Let $\mathbf{S}^{(1)} = \int_0^1 \int_0^1 \partial_x \partial_\sigma \mathbf{S}^{-1}(s_{m-1}, s \hat{X}_m + (1-s)X_{s_{m-1}}, \sigma_t, v_*) ds dt$, $\mathbf{S}^{(2)} = \int_0^1 \partial_x \partial_\sigma \mathbf{S}^{-1}(s_{m-1}, X_{s_{m-2}}, \sigma_t, v_*) dt$ and $\mathbf{S}^{(3)} = \mathbf{S}(s_{m-1}, X_{s_{m-2}}, \sigma_*, v_*)$. Then we obtain

$$\begin{aligned} E_\Pi[|\hat{\Psi}_{2,n}|^q] &= 2^{-q} E_\Pi \left[\left(\sum_m \tilde{Z}_m^\top \mathbf{S}^{(1)} \tilde{Z}_m (\hat{X}_m - X_{s_{m-1}}) (\sigma - \sigma_*) \right)^q \right] \\ &\leq C E_\Pi \left[\left(\sum_m \text{tr}(\mathbf{S}^{(1)} E_m[\tilde{Z}_m \tilde{Z}_m^\top]) (\hat{X}_m - X_{s_{m-1}}) \right)^q \right] \\ &\quad + C E_\Pi \left[\left(\sum_m \text{tr}(\mathbf{S}^{(1)} \bar{E}_m[\tilde{Z}_m \tilde{Z}_m^\top])^2 (\hat{X}_m - X_{s_{m-1}})^2 \right)^{q/2} \right] \\ &\leq C E_\Pi \left[\left(\sum_m \text{tr}(\mathbf{S}^{(2)} \mathbf{S}^{(3)}) (\hat{X}_m - X_{s_{m-1}}) \right)^q \right] + \bar{R}_n((\ell_n b_n^{-1/2} k_n \ell_n^{-1})^q) + \bar{R}_n(\ell_n^{q/2} b_n^{-q/2} k_n^q \ell_n^{-q/2}) \\ &= \bar{R}_n(\ell_n^{q/2} b_n^{-q/2} k_n^q \ell_n^{-q/2}) + \bar{R}_n(\ell_n^q b_n^{-q/2} k_n^q \ell_n^{-q}) + \bar{R}_n(b_n^{-q/2} k_n^q) = \bar{R}_n(b_n^{-q/2} k_n^q) \end{aligned}$$

for any $q > 0$, by the Burkholder–Davis–Gundy inequality, $E_{m-1}[\hat{X}_m - X_{s_{m-1}}] = \bar{R}_n(\ell_n^{-1})$ and $E_\Pi[|\hat{X}_m - X_{s_{m-1}}|^q]^{1/q} = \bar{R}_n(\ell_n^{-1/2})$. Similar estimates for $\partial_\sigma^j \hat{\Psi}_{2,n}$ and Sobolev's inequality yield $E_\Pi[\sup_\sigma |b_n^{-1/2} \partial_\sigma^j \hat{\Psi}_{2,n}|^q] = \bar{R}_n(b_n^{-q} k_n^q)$ for $0 \leq j \leq 3$ and $q > 0$.

Similarly, we have $E_\Pi[\sup_\sigma |b_n^{-1/2} \partial_\sigma^j \hat{\Psi}_{3,n}|^q] = \bar{R}_n(b_n^{-q} k_n^q)$, and therefore we obtain $E_\Pi[(b_n^{-1/2} \sup_\sigma |\partial_\sigma^j (H_n(\sigma, v_*) - H_n(\sigma_*, v_*)) - \partial_\sigma^j (\tilde{H}_n(\sigma, v_*) - \tilde{H}_n(\sigma_*, v_*))|)^q] = \bar{R}_n((b_n^{-5} k_n^7)^{q/4})$ for $0 \leq j \leq 3$.

Taylor's formula yields

$$\begin{aligned}
& H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n) - (H_n(\sigma, v_*) - H_n(\sigma_*, v_*)) \\
&= -\frac{1}{2} \sum_m \int_0^1 \left\{ Z_m^\top (\partial_\sigma S_m^{-1}(\sigma_t, \hat{v}_n) - \partial_\sigma S_m^{-1}(\sigma_t, v_*)) Z_m + \partial_\sigma \log \frac{\det S(\sigma_t, \hat{v}_n)}{\det S(\sigma_t, v_*)} \right\} dt (\sigma - \sigma_*) \\
&= -\frac{1}{2} \sum_m \sum_{j=1}^3 \int_0^1 \left\{ Z_m^\top \partial_v^j \partial_\sigma S_m^{-1}(\sigma_t, v_*) Z_m + \partial_v^j \partial_\sigma \log \det S(\sigma_t, v_*) \right\} dt (\sigma - \sigma_*) \frac{(\hat{v}_n - v_*)^j}{j!} \\
&\quad - \frac{1}{2} \sum_m \int_0^1 \int_0^1 \left\{ Z_m^\top \partial_v^4 \partial_\sigma S_m^{-1}(\sigma_t, v_s) Z_m + \partial_v^4 \partial_\sigma \log \det S(\sigma_t, v_s) \right\} ds dt (\sigma - \sigma_*) \frac{(\hat{v}_n - v_*)^4}{4!},
\end{aligned}$$

where $v_s = s\hat{v}_n + (1-s)v_*$.

Then we obtain

$$\begin{aligned}
& b_n^{-1/4} |H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n) - (H_n(\sigma, v_*) - H_n(\sigma_*, v_*))| \\
&= \bar{R}_n(b_n^{-1/4} b_n^{-1} k_n^2 \ell_n) \times O_p(b_n^{-1/2}) + \bar{R}_n(b_n^{-1/4} (b_n k_n \ell_n)) \times O_p(b_n^{-2}) \rightarrow^p 0,
\end{aligned}$$

by Lemma 4.3, [V], and the equation $\partial_\sigma \log \det S = -\text{tr}(\partial_\sigma S S^{-1})$. Similarly, we obtain

$b_n^{-1/4} |\partial_\sigma^j H_n(\sigma, \hat{v}_n) - \partial_\sigma^j H_n(\sigma, v_*)| \rightarrow^p 0$ for $1 \leq j \leq 4$. Sobolev's inequality yields $\sup_\sigma (b_n^{-1/4} |\partial_\sigma^j (H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n)) - \partial_\sigma^j (H_n(\sigma, v_*) - H_n(\sigma_*, v_*))|) \rightarrow^p 0$ for $0 \leq j \leq 3$ and consequently we obtain $\sup_\sigma (b_n^{-1/2} |\partial_\sigma^j (H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n)) - \partial_\sigma^j (\tilde{H}_n(\sigma, v_*) - \tilde{H}_n(\sigma_*, v_*))|) \rightarrow^p 0$ for $0 \leq j \leq 3$.

Moreover, point 3 of Lemma 4.3 yields $E_\Pi[b_n^{-1/4} \partial_\sigma \hat{\Psi}_{1,n}(\sigma_*)^2] = \bar{R}_n((b_n^{-3/2} k_n^2) \vee (b_n^{-5/2} k_n^{7/2})) \rightarrow^p 0$, and consequently $b_n^{-1/4} (\partial_\sigma H_n(\sigma_*, \hat{v}_n) - \partial_\sigma \tilde{H}_n(\sigma_*, v_*)) \rightarrow^p 0$.

If further [B4] is satisfied, then for any $q > 0$, we obtain

$$\begin{aligned}
& \sup_\sigma E_\Pi[b_n^{-q/2} |H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n) - (H_n(\sigma, v_*) - H_n(\sigma_*, v_*))|^q] \\
&= \bar{R}_n(b_n^{-q/2} b_n^{-q} k_n^{2q} \ell_n^q b_n^{-q/2}) + \bar{R}_n(b_n^{-q/2} (b_n k_n \ell_n)^q b_n^{-2q}) = \bar{R}_n(b_n^{-q} k_n^q),
\end{aligned}$$

by Lemma 4.3, [V], and the equation $\partial_\sigma \log \det S = -\text{tr}(\partial_\sigma S S^{-1})$. Similarly, we obtain

$\sup_\sigma E_\Pi[b_n^{-q/2} |\partial_\sigma^j H_n(\sigma, \hat{v}_n) - \partial_\sigma^j H_n(\sigma, v_*)|^q] = \bar{R}_n(b_n^{-q} k_n^q)$ for $1 \leq j \leq 4$. Sobolev's inequality yields $E_\Pi[\sup_\sigma (b_n^{-1/2} |\partial_\sigma^j (H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n)) - \partial_\sigma^j (H_n(\sigma, v_*) - H_n(\sigma_*, v_*))|)^q] = \bar{R}_n(b_n^{-q} k_n^q)$ for $0 \leq j \leq 3$, which completes the proof. \square

5 The limit of the quasi-likelihood function

We complete the proof of Proposition 2.1 in this section. To do so, it is essential to specify the asymptotic behavior of some functions of approximate covariance matrix \tilde{S}_m , as seen in (5.1). Unlike previous studies by Gloter and Jacod [12, 13], the eigenvalues of the diagonal blocks $\tilde{D}_{1,m}$ and $\tilde{D}_{2,m}$ of \tilde{S}_m are not identified because of the irregular sampling, and even the sizes of $\tilde{D}_{1,m}$ and $\tilde{D}_{2,m}$ are different. These problems make it difficult to deduce asymptotic behaviors of the right-hand side of (5.1). To solve these problems, in Lemma 5.1, we approximate $\tilde{D}_{j,m}$ by $\dot{D}_{j,m}$, which is a kind of local averaged versions of $\tilde{D}_{j,m}$ and has similar properties to the covariance matrix of equidistant sampling scheme. Moreover, we can also change the sizes of $\tilde{D}_{j,m}$ using some specific properties of $\dot{D}_{j,m}$. We deal with this in Lemma 5.2, and show convergence of some trace functions that appear in a decomposition of H_n . The decomposition (4.4) and the nice properties of p_i in Lemma 4.1 are essential in the proofs.

Lemma 4.4 yields

$$\begin{aligned}
& b_n^{-1/4} \partial_\sigma^j H_n(\sigma, \hat{v}_n) \\
&= b_n^{-1/4} \partial_\sigma^j \tilde{H}_n(\sigma, v_*) + o_p(1) \\
&= -\frac{1}{2} b_n^{-\frac{1}{4}} \sum_m (E_m[\tilde{Z}_m^\top \partial_\sigma^j \tilde{S}_m^{-1} \tilde{Z}_m] + \partial_\sigma^j \log \det \tilde{S}_m) - \frac{1}{2} b_n^{-\frac{1}{4}} \sum_m (\tilde{Z}_m^\top \partial_\sigma^j \tilde{S}_m^{-1} \tilde{Z}_m - E_m[\tilde{Z}_m^\top \partial_\sigma^j \tilde{S}_m^{-1} \tilde{Z}_m]) + o_p(1)
\end{aligned}$$

for $1 \leq j \leq 4$. Together with the relation $E_m[\tilde{Z}_m^\top \partial_\sigma^j \tilde{S}_m^{-1} \tilde{Z}_m] = \text{tr}(\partial_\sigma^j \tilde{S}_m^{-1} \tilde{S}_{m,*})$, we obtain

$$b_n^{-\frac{1}{2}} \partial_\sigma^j (H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n)) = -\frac{1}{2} b_n^{-\frac{1}{2}} \sum_m \partial_\sigma^j \left(\text{tr}(\tilde{S}_m^{-1} \tilde{S}_{m,*} - \mathcal{E}) + \log \frac{\det \tilde{S}_m}{\det \tilde{S}_{m,*}} \right) + o_p(1), \quad (5.1)$$

since the residual terms are $o_p(1)$ by Lemma 4.3.

We first investigate asymptotics of $\text{tr}(\tilde{S}_m^{-1} \tilde{S}_{m,*} - \mathcal{E})$. Let $\tilde{L} = \{\tilde{b}_m^1 \cdot \tilde{b}_m^2 | I_{i,m}^1 \cap I_{j,m}^2 | \}_{i,j}$ and $\tilde{G} = \{|I_{i,m}^1 \cap I_{j,m}^2 | \}_{i,j}$. Then since

$$\begin{aligned} \tilde{S}_m^{-1} &= \tilde{D}_m^{-1/2} \sum_{p=0}^{\infty} (-1)^p \begin{pmatrix} 0 & \tilde{D}_{1,m}^{-1/2} \tilde{L} \tilde{D}_{2,m}^{-1/2} \\ \tilde{D}_{2,m}^{-1/2} \tilde{L}^\top \tilde{D}_{1,m}^{-1/2} & 0 \end{pmatrix}^p \tilde{D}_m^{-1/2} \\ &= \sum_{p=0}^{\infty} \begin{pmatrix} \tilde{D}_{1,m}^{-1/2} (\tilde{D}_{1,m}^{-1/2} \tilde{L} \tilde{D}_{2,m}^{-1/2} \tilde{L}^\top \tilde{D}_{1,m}^{-1/2})^p \tilde{D}_{1,m}^{-1/2} & -\tilde{D}_{1,m}^{-1} \tilde{L} \tilde{D}_{2,m}^{-1/2} (\tilde{D}_{2,m}^{-1/2} \tilde{L}^\top \tilde{D}_{1,m}^{-1/2} \tilde{L} \tilde{D}_{2,m}^{-1/2})^p \tilde{D}_{2,m}^{-1/2} \\ -\tilde{D}_{2,m}^{-1} \tilde{L}^\top \tilde{D}_{1,m}^{-1/2} (\tilde{D}_{1,m}^{-1/2} \tilde{L} \tilde{D}_{2,m}^{-1/2} \tilde{L}^\top \tilde{D}_{1,m}^{-1/2})^p \tilde{D}_{1,m}^{-1/2} & \tilde{D}_{2,m}^{-1/2} (\tilde{D}_{2,m}^{-1/2} \tilde{L}^\top \tilde{D}_{1,m}^{-1/2} \tilde{L} \tilde{D}_{2,m}^{-1/2})^p \tilde{D}_{2,m}^{-1/2} \end{pmatrix}, \end{aligned}$$

we have

$$\begin{aligned} &\text{tr}(\tilde{S}_m^{-1} \tilde{S}_{m,*} - \mathcal{E}) \\ &= \sum_{p=0}^{\infty} \{ (|\tilde{b}_{m,*}^1|^2 - |\tilde{b}_m^1|^2) \text{tr}((\tilde{D}_{1,m}^{-1} \tilde{L} \tilde{D}_{2,m}^{-1} \tilde{L}^\top)^p \tilde{D}_{1,m}^{-1} D'_{1,m}) + (|\tilde{b}_{m,*}^2|^2 - |\tilde{b}_m^2|^2) \text{tr}((\tilde{D}_{2,m}^{-1} \tilde{L}^\top \tilde{D}_{1,m}^{-1} \tilde{L})^p \tilde{D}_{2,m}^{-1} D'_{2,m}) \\ &\quad - 2(\tilde{b}_{m,*}^1 \cdot \tilde{b}_{m,*}^2 - \tilde{b}_m^1 \cdot \tilde{b}_m^2) \text{tr}(\tilde{D}_{1,m}^{-1} \tilde{L} \tilde{D}_{2,m}^{-1} (\tilde{L}^\top \tilde{D}_{1,m}^{-1} \tilde{L} \tilde{D}_{2,m}^{-1})^p \tilde{G}^\top) \}. \end{aligned} \quad (5.2)$$

Note that $\|\tilde{G}\| \vee \|\tilde{G}^\top\| \leq r_n$ by Lemma 2 in [26].

We will see the limit of each term on the right-hand side. Let $\hat{a}_m^j = a_{s_{m-1}}^j$ and $\dot{D}_{j,m} = |\tilde{b}_m^j|^2 b_n^{-1} (\hat{a}_m^j)^{-1} \mathcal{E} + v_{j,*} M_{j,m}$. It is difficult to calculate each element or eigenvalue of $\tilde{D}_{j,m}^{-1}$. However, we can apply (4.4) to $\dot{D}_{j,m}^{-1}$, and hence we can calculate its elements. Therefore, we replace $\tilde{D}_{j,m}$ by $\dot{D}_{j,m}$ using the following lemma.

Lemma 5.1. *Let $j \in \{1, 2\}$ and $A_{n,m}$ be a $k_m^j \times k_m^j$ matrix for $1 \leq m \leq \ell_n$. Assume [B1], [A2] and that all elements of $A_{n,m}$ are nonnegative and $\|A_{n,m}\| \leq 1$ for any m . Then*

$$\text{tr}(\partial_\sigma^k \tilde{D}_{j,m}^{-1} A_{n,m}) = \text{tr}(\partial_\sigma^k \dot{D}_{j,m}^{-1} A_{n,m}) + \dot{R}_n(b_n^{3/2} \ell_n^{-1})$$

for $0 \leq k \leq 3$. If further [B2] is satisfied, then

$$\sup_\sigma |\text{tr}(\partial_\sigma^k \tilde{D}_{j,m}^{-1} A_{n,m}) - \text{tr}(\partial_\sigma^k \dot{D}_{j,m}^{-1} A_{n,m})| = \underline{R}_n(b_n^{3/2} \ell_n^{-1})$$

for $0 \leq k \leq 3$.

Proof. We first consider the case where $k = 0$. (4.4), (4.6), and the equation above it yield

$$\begin{aligned} &\text{tr}(\tilde{D}_{j,m}^{-1} A_{n,m}) \\ &= \sum_{p=0}^{\infty} \text{tr} \left(\tilde{D}_{j,m}^{-1/2} (\tilde{D}_{j,m}^{-1/2} (\tilde{D}_{j,m} - \tilde{D}_{j,m}) \tilde{D}_{j,m}^{-1/2})^p \tilde{D}_{j,m}^{-1/2} A_{n,m} \right) \\ &= \frac{1}{v_{j,*}^{p+1}} \sum_{p=0}^{\infty} \sum_{i_1, \dots, i_{p+1}} \frac{1}{p_{i_1} \cdots p_{i_{p+1}} (|\tilde{b}_m^j|^2 r_n v_{j,*}^{-1})} \sum_{\substack{l_q \leq i_q \wedge i_{q+1} \\ l' \leq i_{p+1}, l'' \leq i_1}} \frac{(A_{n,m})_{l', l''}}{P_{l', l'', i_{p+1}, i_1}} \prod_{q=1}^p \frac{(\tilde{D}_{j,m} - \tilde{D}_{j,m})_{l_q, l_q}}{P_{l_q, l_q, i_q, i_{q+1}}}, \end{aligned} \quad (5.3)$$

where $P_{k_1, k_2, l_1, l_2} = \prod_{m_1; k_1 \leq m_1 \leq l_1-1} p_{m_1} \prod_{m_2; k_2 \leq m_2 \leq l_2-1} p_{m_2} (|\tilde{b}_m^j|^2 r_n v_{j,*}^{-1})$.

Then the nice properties of p_i in Lemma 4.1 will lead us to the desired results. Roughly speaking, we have $1 \leq p_i(|\tilde{b}_m^j|^2 r_n v_{j,*}^{-1}) \sim 1 + C b_n^{-1/2}$ for sufficiently large i . This means that P_{k_1, k_2, l_1, l_2} and $P_{k'_1, k'_2, l_1, l_2}$ are asymptotically equivalent if $|k_1 - k'_1|$ and $|k_2 - k'_2|$ are of order less than $b_n^{1/2}$. Then we can replace $\tilde{D}_{j,m}$ in the

right-hand side of (5.3) by $\dot{D}_{j,m}$ since the diagonal elements of both matrices have the same local average. We will verify these rough sketches by the following.

We first see that terms containing small l_q can be ignored. Let η be the one in [A2], $\eta' \in (\eta, 1/2)$, $t_{\tilde{l},m} = s_{m-1} + T[b_n k_n^{-1}]^{-1} b_n^{\eta'} / k_n + T[b_n k_n^{-1}]^{-1} (k_n - b_n^{\eta'}) [(k_n - b_n^{\eta'}) b_n^{-\eta}]^{-1} \tilde{l} / k_n$ for $0 \leq \tilde{l} \leq [(k_n - b_n^{\eta'}) b_n^{-\eta}]$, $\mathcal{I}_m(\tilde{l}) = \{l; I_{l,m}^j \subset [t_{\tilde{l}-1,m}, t_{\tilde{l},m}]\}$, and $\mathcal{E}' = \{\delta_{i_1, i_2} 1_{\{\inf I_{i_1, m}^j < t_0\}}\}_{1 \leq i_1, i_2 \leq k_m^j}$. Then the absolute value of summation involving terms with l_q satisfying $\inf I_{l_q, m}^j < t_0$ for some $1 \leq q \leq p$ on the right-hand side of the above equation is less than

$$\|A_{n,m}\| \sum_{p=1}^{\infty} p \|\check{D}_{j,m}^{-1/2} (\check{D}_{j,m} - \tilde{D}_{j,m}) \check{D}_{j,m}^{-1/2}\|^{p-1} \|\check{D}_{j,m}^{-1/2}\|^2 |\tilde{b}_m^j|^2 \text{tr}(\check{D}_{j,m}^{-1/2} (r_n - \underline{r}_n) \mathcal{E}' \check{D}_{j,m}^{-1/2}) \leq r_n^2 \underline{r}_n^{-2} \text{tr}(\check{D}_{j,m}^{-1} \mathcal{E}'), \quad (5.4)$$

by (4.6), Lemma A.1, and the assumptions. Moreover, point 2 of Lemma 4.1 ensures that $\text{tr}(\check{D}_{j,m}^{-1} \mathcal{E}')$ is less than $[[k_m^j/2]/[T b_n^{-1+\eta'} \underline{r}_n^{-1}/2]]^{-1} \text{tr}(\check{D}_{j,m}^{-1})$ if $[k_m^j/2] \geq 2[T b_n^{-1+\eta'} \underline{r}_n^{-1}]$, and hence the right-hand side of (5.4) is $\bar{R}_n(\mathcal{V}_n)$ by $\text{tr}(\check{D}_{j,m}^{-1}) = \bar{R}_n(r_n^{-1/2} k_n)$, where $\mathcal{V}_n = b_n^{1/2+\eta'} 1_{\{[\bar{k}_n/2] \geq 2[T b_n^{-1+\eta'} \underline{r}_n^{-1}]\}} + b_n^{1/2} k_n 1_{\{[\bar{k}_n/2] < 2[T b_n^{-1+\eta'} \underline{r}_n^{-1}]\}}$.

Then for \tilde{l} , i_q, i_{q+1} and l_q satisfying $l_q \in \mathcal{I}_m(\tilde{l})$ and $\max \mathcal{I}_m(\tilde{l}) \leq i_q \wedge i_{q+1}$, Lemma 4.1 yields that $P_{\max \mathcal{I}_m(\tilde{l}), \max \mathcal{I}_m(\tilde{l}), i_q, i_{q+1}} / P_{l_q, l_q, i_q, i_{q+1}}$ is less than 1 and greater than

$$(1 + C^{-1} b_n^{1-\eta'} \underline{r}_n + C |\tilde{b}_m^j|^2 b_n^{-1+\eta'} r_n \underline{r}_n^{-1})^{-C b_n^{-1+\eta} \underline{r}_n^{-1}} \geq 1 - \bar{R}_n(b_n^{\eta-\eta'} + b_n^{-2+\eta'+\eta} \underline{r}_n^{-2} r_n)$$

for sufficiently large n . Moreover, $\max_{\tilde{l}} |\sum_{l \in \mathcal{I}_m(\tilde{l})} (\check{D}_{j,m} - \dot{D}_{j,m})_{l,l}|$ is $\dot{R}_n(b_n^{-1+\eta})$ by [A2]. We also have $\sup_{\sigma, m} \max_{\tilde{l}} |\sum_{l \in \mathcal{I}_m(\tilde{l})} (\check{D}_{j,m} - \dot{D}_{j,m})_{l,l}| = \underline{R}_n(b_n^{-1+\eta})$ if [B2] is satisfied.

Therefore we obtain

$$\begin{aligned} & \text{tr}(\check{D}_{j,m}^{-1} A_{n,m}) \\ &= \sum_{p=0}^{\infty} \sum_{i_1, \dots, i_{p+1}} \frac{1}{p_{i_1} \cdots p_{i_{p+1}} (|\tilde{b}_m^j|^2 r_n)} \sum_{\substack{l_q \leq i_q \wedge i_{q+1}, \inf I_{l_q, m}^j \geq t_0 \\ l' \leq i_{p+1}, l'' \leq i_1}} \frac{(A_{n,m})_{l', l''}}{P_{l', l'', i_{p+1}, i_1}} \prod_{q=1}^p \frac{(\check{D}_{j,m} - \tilde{D}_{j,m})_{l_q, l_q}}{P_{l_q, l_q, i_q, i_{q+1}}} + \bar{R}_n(\mathcal{V}_n) \\ &= \sum_{p=0}^{\infty} \mathcal{T}_{m,1}^{n,p} \sum_{i_1, \dots, i_{p+1}} \frac{1}{p_{i_1} \cdots p_{i_{p+1}} (|\tilde{b}_m^j|^2 r_n)} \sum_{l' \leq i_{p+1}, l'' \leq i_1} \frac{(A_{n,m})_{l', l''}}{P_{l', l'', i_{p+1}, i_1}} \\ & \quad \times \prod_{q=1}^p \sum_{1 \leq \tilde{l}_q \leq [b_n^{\eta}]^{-1} (k_n - [b_n^{\eta'}])} \frac{\sum_{l_q \in \mathcal{I}_m(\tilde{l}_q)} (\check{D}_{j,m} - \dot{D}_{j,m} + \mathcal{T}_{m,3}^{n,p} \mathcal{E})_{l_q, l_q}}{P_{\max \mathcal{I}_m(\tilde{l}_q), \max \mathcal{I}_m(\tilde{l}_q), i_q, i_{q+1}}} + \bar{R}_n(\mathcal{V}_n) \\ &= \sum_{p=0}^{\infty} \mathcal{T}_{m,1}^{n,p} (\mathcal{T}_{m,2}^{n,p})^{-1} \text{tr}(\check{D}_{j,m}^{-1/2} (\check{D}_{j,m}^{-1/2} (\check{D}_{j,m} - \dot{D}_{j,m} + \mathcal{T}_{m,3}^{n,p} \mathcal{E}) \check{D}_{j,m}^{-1/2})^p \check{D}_{j,m}^{-1/2} A_{n,m}) + \dot{R}_n(b_n^{3/2} \ell_n^{-1}), \quad (5.5) \end{aligned}$$

where $\mathcal{T}_{m,i}^{n,p}$ is a random variable which does not depend on $l_q, \tilde{l}_q, i_q, i_{q+1}$ and satisfies

$$(1 - \bar{R}_n(b_n^{\eta-\eta'} + b_n^{-2+\eta'+\eta} \underline{r}_n^{-2} r_n))^p \leq \mathcal{T}_{m,i}^{n,p} \leq 1$$

for $i = 1, 2$ and $\sup_p |\mathcal{T}_{m,3}^{n,p}| = \dot{R}_n(b_n^{-1})$.

Let $F_p(t) = \text{tr}(\check{D}_{j,m}^{-1/2} (\check{D}_{j,m}^{-1/2} (\check{D}_{j,m} - \dot{D}_{j,m} + t \mathcal{T}_{m,3}^{n,p} \mathcal{E}) \check{D}_{j,m}^{-1/2})^p \check{D}_{j,m}^{-1/2} A_{n,m})$. Then

$$|F_p(1) - F_p(0)| \leq \int_0^1 |F_p'(t)| dt \leq p |\tilde{b}_m^j|^{-2} |\mathcal{T}_{m,3}^{n,p}| (1 - b_n^{-1} (\hat{a}_m^j)^{-1} r_n^{-1} + |\tilde{b}_m^j|^{-2} \mathcal{T}_{m,3}^{n,p} r_n^{-1})^{p-1} \bar{R}_n(r_n^{-1/2} b_n^2 \ell_n^{-1}),$$

and hence $\sum_{p=0}^{\infty} |\mathcal{T}_{m,1}^{n,p} (\mathcal{T}_{m,2}^{n,p})^{-1}| |F_p(1) - F_p(0)| \leq \sup_p |\mathcal{T}_{m,3}^{n,p}| \cdot \dot{R}_n(b_n^{3/2} k_n) = \dot{R}_n(b_n^{3/2} \ell_n^{-1})$. Therefore we obtain

the desired conclusion by

$$\begin{aligned} & \sum_{p=0}^{\infty} |\mathcal{T}_{m,1}^{n,p}(\mathcal{T}_{m,2}^{n,p})^{-1} - 1| F_p(0) \\ & \leq C \sum_p p \bar{R}_n(b_n^{\eta-\eta'} + b_n^{-1+\eta+\eta'}) (1 + \bar{R}_n(b_n^{\eta-\eta'} + b_n^{-1+\eta+\eta'}))^{p-1} (1 - b_n^{-1}(\hat{a}_m^j)^{-1} r_n^{-1})^p r_n^{-1/2} k_m^j = \dot{R}_n(b_n^{3/2} \ell_n^{-1}). \end{aligned}$$

For the case $k = 1$, we have $\text{tr}(\partial_\sigma \tilde{D}_{j,m}^{-1} A_{n,m}) = \text{tr}(\dot{D}_{j,m}^{-1} \partial_\sigma \tilde{D}_{j,m} \dot{D}_{j,m}^{-1} A_{n,m}) + \dot{R}_n(b_n^{3/2} \ell_n^{-1})$ by using the result for $k = 0$, $\partial_\sigma \tilde{D}_{j,m}^{-1} = \tilde{D}_{j,m}^{-1} \partial_\sigma \tilde{D}_{j,m} \tilde{D}_{j,m}^{-1}$, $\|\tilde{D}_{j,m}^{-1} \partial_\sigma \tilde{D}_{j,m}\| = \bar{R}_n(1)$ and all elements of $\tilde{D}_{j,m}^{-1}$ are nonnegative by a similar argument to (4.4). Then a similar argument to (5.5) enables us to replace $\partial_\sigma \tilde{D}_{j,m}$ by $\partial_\sigma \dot{D}_{j,m}$. Similarly, we obtain $\text{tr}(\partial_\sigma^k \tilde{D}_{j,m}^{-1} A_{n,m}) = \text{tr}(\partial_\sigma^k \dot{D}_{j,m}^{-1} A_{n,m}) + \dot{R}_n(b_n^{3/2} \ell_n^{-1})$ for $k = 2, 3$.

If further [B2] is satisfied, then similarly we have $\sup_\sigma |\text{tr}(\partial_\sigma^k \tilde{D}_{j,m}^{-1} A_{n,m}) - \text{tr}(\partial_\sigma^k \dot{D}_{j,m}^{-1} A_{n,m})| = \underline{R}_n(b_n^{3/2} \ell_n^{-1})$. \square

Remark 5.1. The proof shows that there are upperbounds of the absolute values of residual terms in the statement of Lemma 5.1 which do not depend on $A_{n,m}$.

Let $c_j = |\tilde{b}_m^j|^2 b_n^{-1} (\hat{a}_m^j)^{-1} / v_{j,*}$ and $c'_j = c_j (\hat{a}_m^j / \hat{a}_m^{3-j})^2$ for $j = 1, 2$.

Lemma 5.2. Assume [B1] and [A2]. Then

$$\sup_{\sigma,m} \left| \partial_\sigma^k \text{tr}((\tilde{D}_{1,m}^{-1} \tilde{G} \tilde{D}_{2,m}^{-1} \tilde{G}^\top)^p \tilde{D}_{1,m}^{-1} D'_{1,m}) - \frac{T \hat{a}_m^1 k_n}{\pi} \frac{(\hat{a}_m^2)^p \partial_\sigma^k I_{p+1,p}(c_1, c'_2)}{b_n^{2p+1} (\hat{a}_m^1)^{3p+1} v_{1,*}^p v_{2,*}^p} \right| = o_p(b_n^{1/2} \ell_n^{-1}), \quad (5.6)$$

$$\sup_{\sigma,m} \left| \partial_\sigma^k \text{tr}((\tilde{D}_{2,m}^{-1} \tilde{G}^\top \tilde{D}_{1,m}^{-1} \tilde{G})^p \tilde{D}_{2,m}^{-1} D'_{2,m}) - \frac{T \hat{a}_m^1 k_n}{\pi} \frac{(\hat{a}_m^2)^{p+1} \partial_\sigma^k I_{p,p+1}(c_1, c'_2)}{b_n^{2p+1} (\hat{a}_m^1)^{3p+2} v_{1,*}^p v_{2,*}^{p+1}} \right| = o_p(b_n^{1/2} \ell_n^{-1}), \quad (5.7)$$

$$\sup_{\sigma,m} \left| \partial_\sigma^k \text{tr}((\tilde{D}_{1,m}^{-1} \tilde{G} \tilde{D}_{2,m}^{-1} \tilde{G}^\top)^{p+1}) - \frac{T \hat{a}_m^1 k_n}{\pi} \frac{(\hat{a}_m^2)^{p+1} \partial_\sigma^k I_{p+1,p+1}(c_1, c'_2)}{(b_n^2 (\hat{a}_m^1)^3 v_{1,*} v_{2,*})^{p+1}} \right| = o_p(b_n^{1/2} \ell_n^{-1}) \quad (5.8)$$

for $0 \leq k \leq 3$ and $p \in \mathbb{Z}_+$. If further [B2] is satisfied, then $o_p(b_n^{1/2} \ell_n^{-1})$ in (5.6)-(5.8) can be replaced by $\underline{R}_n(b_n^{1/2} \ell_n^{-1})$.

Proof. For any $p \in \mathbb{N}$, Lemma 5.1 yields

$$b_n^{-1/2} \partial_\sigma^k \text{tr}((\tilde{D}_{1,m}^{-1} \tilde{G} \tilde{D}_{2,m}^{-1} \tilde{G}^\top)^p) = b_n^{-1/2} \partial_\sigma^k \text{tr}((\dot{D}_{1,m}^{-1} \tilde{G} \dot{D}_{2,m}^{-1} \tilde{G}^\top)^p) + \dot{R}_n(\ell_n^{-1}). \quad (5.9)$$

Moreover, we have

$$b_n^{-1/2} \text{tr}((\dot{D}_{1,m}^{-1} \tilde{G} \dot{D}_{2,m}^{-1} \tilde{G}^\top)^p) = b_n^{-1/2} \frac{1}{v_{1,*}^p v_{2,*}^p} \sum_{\substack{i_1, \dots, i_p \\ j_1, \dots, j_p}} \sum_{\substack{\alpha_{2q-1} \leq i_q, \beta_{2q-1} \leq j_q \\ \alpha_{2q} \leq i_{q+1}, \beta_{2q} \leq j_{q+1} \\ (1 \leq q \leq p)}} \prod_{q=1}^p \frac{\tilde{G}_{\alpha_{2q-1}, \beta_{2q-1}}}{\tilde{P}_{\alpha_{2q-1}, \beta_{2q-1}, i_q+1, j_q+1}} \frac{\tilde{G}_{\alpha_{2q}, \beta_{2q}}}{\tilde{P}_{\alpha_{2q}, \beta_{2q}, i_{q+1}, j_{q+1}}} \quad (5.10)$$

by (4.4), where $i_{p+1} = i_1$ and $\tilde{P}_{\alpha, \beta, i, j} = \prod_{k_1=\alpha}^{i-1} p_{k_1}(c_1) \prod_{k_2=\beta}^{j-1} p_{k_2}(c_2)$.

We will apply (4.2) to obtain the limit of the traces. To do so, we need to change the size of matrices \tilde{G} and $\dot{D}_{2,m}^{-1}$. This is again achieved by the nice properties of p_i . The essential idea is that point 3 of Lemma 4.1 ensures $p_i \sim p_+$ for sufficiently large i , and therefore $\tilde{P}_{\alpha, \beta, i, j} \sim p_+(c_1)^{i-\alpha} p_+(c_2)^{j-\beta} \sim \exp(\sqrt{c_1}(i-\alpha) + \sqrt{c_2}(j-\beta)) \sim \dot{P}_{k\alpha, k\beta, ki, kj}$, where $k \in \mathbb{N}$ and $\dot{P}_{\alpha', \beta', i', j'} = \prod_{k_1=\alpha'}^{i'-1} p_{k_1}(c_1/k^2) \prod_{k_2=\beta'}^{j'-1} p_{k_2}(c_2/k^2)$. The size of $\dot{D}_{2,m}^{-1}$ decides the ranges of summation of j_1, \dots, j_p in (5.10). By changing these ranges using the above relation on $\tilde{P}_{\alpha, \beta, i, j}$ and $\dot{P}_{k\alpha, k\beta, ki, kj}$, we can change the size of matrices \tilde{G} and $\dot{D}_{2,m}^{-1}$.

Now we verify the above idea. First, we see that the terms involving small α_q or β_q in (5.10) can be ignored. Let $\eta \in (0, 1/2)$ be the one in [A2], $\delta \in (1/2, 1)$ such that $b_n^\delta k_n^{-1} \rightarrow 0$, $\tilde{s}_{l'} = s_{m-1} + T[b_n k_n^{-1}]^{-1} [k_n b_n^{-\eta}]^{-1} ((l' + [b_n^{\delta-\eta}]) \wedge [k_n b_n^{-\eta}])$ for $0 \leq l' \leq ([k_n b_n^{-\eta}] - [b_n^{\delta-\eta}]) \vee 0$, $\dot{D}_{3,m} = (c_1 \wedge c_2)(v_{1,*} \wedge v_{2,*}) \mathcal{E}_{k_m^1 \vee k_m^2} + (v_{1,*} \wedge v_{2,*}) M(k_m^1 \vee k_m^2)$,

$G' = \{|I_{i,m}^1 \cap I_{j,m}^2| 1_{\{\inf I_{i,m}^1 \wedge \inf I_{j,m}^2 < \tilde{s}_0\}}\}_{1 \leq i,j \leq k_m^1 \vee k_m^2}$, $\hat{G} = \{|I_{i,m}^1 \cap I_{j,m}^2| 1_{\{i \leq k_m^1 \text{ and } j \leq k_m^2\}}\}_{1 \leq i,j \leq k_m^1 \vee k_m^2}$, and $\mathcal{E}'' = \{\delta_{ij} 1_{\{\inf I_{i,m}^1 \wedge \inf I_{j,m}^2 < \tilde{s}_0\}}\}_{1 \leq i,j \leq k_m^1 \vee k_m^2}$. Similarly to the proof of Lemma 5.1, the absolute value Λ_1 of a summation involving the terms with (α_q, β_q) satisfying $\inf I_{\alpha_q,m}^1 \wedge \inf I_{\beta_q,m}^2 < \tilde{s}_0$ is less than $pb_n^{-1/2} \text{tr}(\dot{D}_{3,m}^{-1}(G' \dot{D}_{3,m}^{-1} \hat{G}^\top + \hat{G} \dot{D}_{3,m}^{-1}(G')^\top)(\dot{D}_{3,m}^{-1} \hat{G} \dot{D}_{3,m}^{-1} \hat{G}^\top)^{p-1})$. Lemma 3 in [26] implies $\|G' + (G')^\top\| \leq 2r_n$ and hence all the eigenvalues of $G' + (G')^\top$ are greater than or equal to $-2r_n$. Therefore $G' + (G')^\top + 2r_n \mathcal{E}''$ is nonnegative definite, and hence Lemma A.1 yields

$$\Lambda_1 \leq p(r_n/\underline{r}_n)^{2p-1} b_n^{-1/2} \text{tr}(\dot{D}_{3,m}^{-1}(G' + (G')^\top + 2r_n \mathcal{E}'')) \leq 2p(r_n/\underline{r}_n)^{2p-1} b_n^{-1/2} (\text{tr}(\dot{D}_{3,m}^{-1} G') + r_n \text{tr}(\dot{D}_{3,m}^{-1} \mathcal{E}'')).$$

Let $(\dot{G})_{i,j} = (\sum_{l \leq i} G'_{l,i} + \sum_{m < i} G'_{i,m}) \delta_{i,j}$ and $\dot{k} = \max\{i; \dot{G}_{ii} > 0\}$. Then Lemma 4.1 yields

$$\begin{aligned} \text{tr}(\dot{D}_{3,m}^{-1} G') &= \frac{1}{v_{1,*} \wedge v_{2,*}} \sum_i \sum_{\alpha, \beta \leq i} \frac{G'_{\alpha,\beta}}{p_\alpha \cdots p_i p_\beta \cdots p_{i-1}} \leq \frac{1}{v_{1,*} \wedge v_{2,*}} \sum_i \sum_{\alpha \leq i} \frac{\dot{G}_{\alpha,\alpha}}{p_\alpha \cdots p_i p_\alpha \cdots p_{i-1}} \\ &= \text{tr}(\dot{D}_{3,m}^{-1} \dot{G}) \leq (\dot{D}_{3,m}^{-1})_{\dot{k},\dot{k}} (\tilde{s}_0 - s_{m-1} + r_n) \leq \frac{C b_n^{-1+\delta} + r_n}{[(k_m^1 \vee k_m^2)/2] - \dot{k}} \text{tr}(\dot{D}_{3,m}^{-1}). \end{aligned}$$

Therefore we obtain $\Lambda_1 = \dot{R}_n(\ell_n^{-1})$, and $\Lambda_1 = \underline{R}_n(\ell_n^{-1})$ if $[B2]$ is satisfied.

Let $\ddot{D}_{2,m} = v_{2,*} c_2 (\hat{a}_m^2 / \hat{a}_m^1)^{2\mathcal{E}} + v_{2,*} M_{1,m}$, $i(\alpha') = \min\{i; S_i^{n,1} \geq \tilde{s}_{\alpha'-1}\}$, and $j(\alpha') = \min\{j; S_j^{n,2} \geq \tilde{s}_{\alpha'-1}\}$. We will show that $b_n^{-1/2} \text{tr}((\dot{D}_{1,m}^{-1} \tilde{G} \dot{D}_{2,m}^{-1} \tilde{G}^\top)^p)$ is approximated by $b_n^{-1/2-2p} (\hat{a}_m^2)^p (\hat{a}_m^1)^{-3p} \text{tr}((\dot{D}_{1,m}^{-1} \ddot{D}_{2,m})^p)$. A similar argument to the proof of Lemma 5.1 yields $|\tilde{P}_{\alpha,\beta,i,j} / \tilde{P}_{i(\alpha'),j(\alpha'),i(i'),j(j')} - 1| = \dot{R}_n(1)$ for $i(\alpha') \leq \alpha < i(\alpha' + 1)$, $j(\alpha') \leq \beta < j(\alpha' + 1)$, $i(i') \leq i < i(i' + 1)$, and $j(j') \leq j < j(j' + 1)$. Therefore repeated use of [A2] yields

$$\begin{aligned} &b_n^{-1/2} \text{tr}((\dot{D}_{1,m}^{-1} \tilde{G} \dot{D}_{2,m}^{-1} \tilde{G}^\top)^p) \\ &= b_n^{-1/2} \frac{\mathcal{T}_{m,4}^{n,p}}{v_{1,*}^p v_{2,*}^p} \sum_{\substack{i'_1, \dots, i'_p \\ j'_1, \dots, j'_p}} \sum_{\substack{\alpha'_{2q-1} \leq i'_q \wedge j'_q \\ \alpha'_{2q} \leq i'_{q+1} \wedge j'_q \quad (1 \leq q \leq p)}} \prod_{q=1}^p \frac{(T b_n^{-1+\eta})^2 \#\{i_q; I_{i_q,m}^1 \subset [\tilde{s}_{i'_q-1}, \tilde{s}_{i'_q}]\} \#\{j_q; I_{j_q,m}^2 \subset [\tilde{s}_{j'_q-1}, \tilde{s}_{j'_q}]\}}{\tilde{P}_{i(\alpha'_{2q-1}),j(\alpha'_{2q-1}),i(i'_q),j(j'_q)} \tilde{P}_{i(\alpha'_{2q}),j(\alpha'_{2q}),i(i'_{q+1}),j(j'_q)}} \\ &\quad + \dot{R}_n(\ell_n^{-1}) \\ &= b_n^{-1/2} \frac{\mathcal{T}_{m,5}^{n,p} (T b_n^{-1+\eta})^{2p} (\hat{a}_m^2)^p}{v_{1,*}^p v_{2,*}^p (\hat{a}_m^1)^p} \\ &\quad \times \sum_{\substack{i'_1, \dots, i'_p \\ j'_1, \dots, j'_p}} \sum_{\substack{\alpha'_{2q-1} \leq i'_q \wedge j'_q \\ \alpha'_{2q} \leq i'_{q+1} \wedge j'_q \quad (1 \leq q \leq p)}} \prod_{q=1}^p \frac{\#\{\alpha; I_{\alpha,m}^1 \subset [\tilde{s}_{\alpha'_{2q-1}-1}, \tilde{s}_{\alpha'_{2q-1}}]\} \#\{\alpha; I_{\alpha,m}^1 \subset [\tilde{s}_{\alpha'_{2q}-1}, \tilde{s}_{\alpha'_{2q}}]\}}{\tilde{P}_{i(\alpha'_{2q-1}),j(\alpha'_{2q-1}),i(i'_q),j(j'_q)} \tilde{P}_{i(\alpha'_{2q}),j(\alpha'_{2q}),i(i'_{q+1}),j(j'_q)}} + \dot{R}_n(\ell_n^{-1}), \end{aligned} \tag{5.11}$$

where $\mathcal{T}_{m,i}^{n,p}$ is a random variable satisfying $\sup_{\sigma,m} |\mathcal{T}_{m,i}^{n,p} - 1| = \dot{R}_n(1)$ for $i = 4, 5$.

Since Lemma 4.1 and [A2] yield

$$\begin{aligned} p_{j(\alpha)} \cdots p_{j(\beta)-1}(c_2) &= (p_+(c_2))^{j(\beta)-j(\alpha)} (1 + \dot{R}_n(1)) = \exp((b_n \hat{a}_m^2 (\tilde{s}_\beta - \tilde{s}_\alpha) + \dot{R}_n(b_n^{1/2})) \log p_+(c_2)) (1 + \dot{R}_n(1)) \\ &= \exp(\hat{a}_m^2 (\hat{a}_m^1)^{-1} (i(\beta) - i(\alpha)) \log p_+(c_2)) (1 + \dot{R}_n(1)) = p_{i(\alpha)} \cdots p_{i(\beta)-1}(c'_2) (1 + \dot{R}_n(1)), \end{aligned}$$

we may replace $\tilde{P}_{i(\alpha'_{2q-1}),j(\alpha'_{2q-1}),i(i'_q),j(j'_q)}$ and $\tilde{P}_{i(\alpha'_{2q}),j(\alpha'_{2q}),i(i'_{q+1}),j(j'_q)}$ in the right-hand side of (5.11) by $\hat{P}_{i(\alpha'_{2q-1}),i(\alpha'_{2q-1}),i(i'_q),j(j'_q)}$ and $\hat{P}_{i(\alpha'_{2q}),i(\alpha'_{2q}),i(i'_{q+1}),i(j'_q)}$, respectively, where $\hat{P}_{\alpha,\beta,i,j} = \prod_{k_1=\alpha}^{i-1} p_{k_1}(c_1) \prod_{k_2=\beta}^{j-1} p_{k_2}(c'_2)$. Therefore, we obtain

$$\sup_{\sigma,m} \left| b_n^{-1/2} \text{tr}((\dot{D}_{1,m}^{-1} \tilde{G} \dot{D}_{2,m}^{-1} \tilde{G}^\top)^p) - b_n^{-1/2-2p} (\hat{a}_m^2)^p (\hat{a}_m^1)^{-3p} \text{tr}((\dot{D}_{1,m}^{-1} \ddot{D}_{2,m})^p) \right| = o_p(\ell_n^{-1}),$$

by a similar argument to (5.11). Since $\partial_\sigma^l \dot{D}_{j,m} = \partial_\sigma^l c_j v_{j,*} \mathcal{E}$ for $1 \leq l \leq 3$, we similarly obtain

$$\sup_{\sigma,m} \left| b_n^{-1/2} \partial_\sigma^k \text{tr}((\dot{D}_{1,m}^{-1} \tilde{G} \dot{D}_{2,m}^{-1} \tilde{G}^\top)^p) - b_n^{-1/2-2p} (\hat{a}_m^2)^p (\hat{a}_m^1)^{-3p} \partial_\sigma^k \text{tr}((\dot{D}_{1,m}^{-1} \ddot{D}_{2,m})^p) \right| = o_p(\ell_n^{-1}). \tag{5.12}$$

Then (4.3), (5.9) and (5.12) yield (5.8).

We also have (5.6) and (5.7) by a similar argument.

Similar arguments enable us to replace $o_p(b_n^{1/2}\ell_n^{-1})$ by $\underline{R}_n(b_n^{1/2}\ell_n^{-1})$ in (5.6)-(5.8) if [B2] is satisfied. \square

Proof of Proposition 2.1.

We first prove the results under the additional condition [A1'].

Since $\|\tilde{D}_{1,m}^{-1/2}\tilde{G}\tilde{D}_{2,m}^{-1}\tilde{G}^\top\tilde{D}_{1,m}^{-1/2}\| \leq |\tilde{b}_m^1|^{-2}|\tilde{b}_m^2|^{-2}$ by Lemma 3 in [26], for any $\epsilon, \delta > 0$, there exists $P_1 \in \mathbb{N}$ such that

$$\begin{aligned} & \sup_n P \left[\sup_\sigma b_n^{-\frac{1}{2}} \sum_m \sum_{p=P+1}^\infty |\partial_\sigma^k((\tilde{b}_m^1 \cdot \tilde{b}_m^2)^{2p-1} \text{tr}((\tilde{D}_{1,m}^{-1}\tilde{G}\tilde{D}_{2,m}^{-1}\tilde{G}^\top)^p))| \geq \delta \right] < \epsilon, \\ & \sup_n P \left[\sup_\sigma b_n^{-\frac{1}{2}} \sum_m \sum_{p=P+1}^\infty \left| \partial_\sigma^k \left(\frac{(\tilde{b}_m^1 \cdot \tilde{b}_m^2)^{2p-1} \hat{a}_m^1 k_n (\hat{a}_m^2)^p I_{p,p}(c_1, c_2')}{(b_n^2 (\hat{a}_m^1)^3 v_{1,*} v_{2,*})^p} \right) \right| \geq \delta \right] < \epsilon \end{aligned}$$

for $P \geq P_1$. Together with Lemma 5.2, we obtain

$$\sup_\sigma \left| b_n^{-1/2} \partial_\sigma^k \sum_m \sum_{p=1}^\infty (\tilde{b}_m^1 \cdot \tilde{b}_m^2)^{2p-1} \text{tr}((\tilde{D}_{1,m}^{-1}\tilde{G}\tilde{D}_{2,m}^{-1}\tilde{G}^\top)^p) - \frac{T \hat{a}_m^1 k_n}{\pi b_n^{1/2}} \partial_\sigma^k \sum_m \sum_{p=1}^\infty \frac{(\tilde{b}_m^1 \cdot \tilde{b}_m^2)^{2p-1} (\hat{a}_m^2)^p I_{p,p}(c_1, c_2')}{(b_n^2 (\hat{a}_m^1)^3 v_{1,*} v_{2,*})^p} \right| \rightarrow^p 0.$$

Let $\dot{a}_m^j = \tilde{a}_{s_{m-1}}^j$, $\mathfrak{C}_m = |\tilde{b}_m^1|^2 |\tilde{b}_m^2|^2 - (\tilde{b}_m^1 \cdot \tilde{b}_m^2)^2$, $\mathfrak{A}_t = \varphi(\tilde{a}_t^1 |b_t^1|^2 + \tilde{a}_t^2 |b_t^2|^2, \tilde{a}_t^1 \tilde{a}_t^2 \det(b_t b_t^\top))$, and

$$P_n = \varphi(c_1 + c_2', b_n^{-2} \hat{a}_m^2 (\hat{a}_m^1)^{-3} v_{1,*}^{-1} v_{2,*}^{-1} \mathfrak{C}_m) = b_n^{-1/2} (\hat{a}_m^1)^{-1} \varphi(\dot{a}_m^1 |\tilde{b}_m^1|^2 + \dot{a}_m^2 |\tilde{b}_m^2|^2, \dot{a}_m^1 \dot{a}_m^2 \mathfrak{C}_m).$$

Then Lemma A.9 yields

$$\begin{aligned} & \frac{T \hat{a}_m^1 k_n}{\pi b_n^{1/2}} \partial_\sigma^k \sum_m \sum_{p=1}^\infty \frac{(\tilde{b}_m^1 \cdot \tilde{b}_m^2)^{2p-1} (\hat{a}_m^2)^p I_{p,p}(c_1, c_2')}{(b_n^2 (\hat{a}_m^1)^3 v_{1,*} v_{2,*})^p} \\ &= \partial_\sigma^k \sum_m \frac{T b_n^{1/2} \hat{a}_m^1 \ell_n^{-1} \frac{\hat{a}_m^2 \tilde{b}_m^1 \cdot \tilde{b}_m^2}{(\hat{a}_m^1)^3 b_n^2 v_{1,*} v_{2,*}}}{b_n^{-1/2} (\hat{a}_m^1)^{-1} \sqrt{2} \varphi(\dot{a}_m^1 |\tilde{b}_m^1|^2 + \dot{a}_m^2 |\tilde{b}_m^2|^2, \dot{a}_m^1 \dot{a}_m^2 \mathfrak{C}_m) \sqrt{\dot{a}_m^1 \dot{a}_m^2} \sqrt{\mathfrak{C}_m} b_n^{-1} (\hat{a}_m^1)^{-2}} + o_p(1) \\ &= \partial_\sigma^k \sum_m \frac{T \ell_n^{-1} \sqrt{\dot{a}_m^1 \dot{a}_m^2} \tilde{b}_m^1 \cdot \tilde{b}_m^2}{\sqrt{2} \mathfrak{C}_m \varphi(\dot{a}_m^1 |\tilde{b}_m^1|^2 + \dot{a}_m^2 |\tilde{b}_m^2|^2, \dot{a}_m^1 \dot{a}_m^2 \mathfrak{C}_m)} + o_p(1) = \partial_\sigma^k \int_0^T \frac{\sqrt{\tilde{a}_t^1 \tilde{a}_t^2} b_t^1 \cdot b_t^2}{\sqrt{2} \det(b_t b_t^\top) \mathfrak{A}_t} dt + o_p(1). \end{aligned}$$

Therefore, we have

$$\sup_\sigma \left| b_n^{-1/2} \partial_\sigma^k \sum_m \sum_{p=1}^\infty (\tilde{b}_m^1 \cdot \tilde{b}_m^2)^{2p-1} \text{tr}((\tilde{D}_{1,m}^{-1}\tilde{G}\tilde{D}_{2,m}^{-1}\tilde{G}^\top)^p) - \partial_\sigma^k \int_0^T \frac{\sqrt{\tilde{a}_t^1 \tilde{a}_t^2} b_t^1 \cdot b_t^2}{\sqrt{2} \det(b_t b_t^\top) \mathfrak{A}_t} dt \right| \rightarrow^p 0. \quad (5.13)$$

Similarly, we obtain

$$\sup_\sigma \left| b_n^{-1/2} \partial_\sigma^k \sum_m \sum_{p=1}^\infty \text{tr}((\tilde{D}_{1,m}^{-1} \tilde{L} \tilde{D}_{2,m}^{-1} \tilde{L}^\top)^p \tilde{D}_{1,m}^{-1} D'_{1,m}) - \partial_\sigma^k \int_0^T \frac{|b_t^2|^2 \sqrt{\tilde{a}_t^1 \tilde{a}_t^2} + \tilde{a}_t^1 \sqrt{\det(b_t b_t^\top)}}{\sqrt{2} \det(b_t b_t^\top) \mathfrak{A}_t} dt \right| \rightarrow^p 0, \quad (5.14)$$

$$\sup_\sigma \left| b_n^{-1/2} \partial_\sigma^k \sum_m \sum_{p=1}^\infty \text{tr}((\tilde{D}_{2,m}^{-1} \tilde{L}^\top \tilde{D}_{1,m}^{-1} \tilde{L})^p \tilde{D}_{2,m}^{-1} D'_{2,m}) - \partial_\sigma^k \int_0^T \frac{|b_t^1|^2 \sqrt{\tilde{a}_t^1 \tilde{a}_t^2} + \tilde{a}_t^2 \sqrt{\det(b_t b_t^\top)}}{\sqrt{2} \det(b_t b_t^\top) \mathfrak{A}_t} dt \right| \rightarrow^p 0. \quad (5.15)$$

Furthermore, Lemma A.3 and a similar argument yield

$$\begin{aligned} \partial_\sigma^k \log \det(\tilde{S}_m \tilde{D}_m^{-1}) &= \partial_\sigma^k \log \det \left(\mathcal{E} + \begin{pmatrix} 0 & \tilde{D}_{1,m}^{-1/2} \tilde{L} \tilde{D}_{2,m}^{-1/2} \\ \tilde{D}_{2,m}^{-1/2} \tilde{L}^\top \tilde{D}_{1,m}^{-1/2} & 0 \end{pmatrix} \right) \\ &= - \sum_{p=1}^\infty \frac{1}{p} \partial_\sigma^k \text{tr}((\tilde{D}_{1,m}^{-1} \tilde{L} \tilde{D}_{2,m}^{-1} \tilde{L}^\top)^p) \\ &= - \frac{T \hat{a}_m^1 k_n}{\pi} \sum_{p=1}^\infty \frac{(\hat{a}_m^2)^p (\tilde{b}_m^1 \cdot \tilde{b}_m^2)^{2p} I_{p,p}(c_1, c_2')}{p (\hat{a}_m^1)^{3p} b_n^{2p} v_{1,*}^p v_{2,*}^p} + o_p(b_n^{1/2} \ell_n^{-1}). \end{aligned}$$

Then Lemma A.9 yields

$$\begin{aligned}
& \partial_\sigma^k \log \det(\tilde{S}_m \tilde{D}_m^{-1}) \\
&= -T \hat{a}_m^1 k_n \left(\sqrt{\frac{|\tilde{b}_m^1|^2}{b_n \hat{a}_m^1 v_{1,*}}} + \sqrt{\frac{\hat{a}_m^2 |\tilde{b}_m^2|^2}{b_n (\hat{a}_m^1)^2 v_{2,*}}} \right) + \frac{T \hat{a}_m^1 k_n b_n^{-1/2}}{\sqrt{2} \hat{a}_m^1} \varphi(\dot{a}_m^1 |\tilde{b}_m^1|^2 + \dot{a}_m^2 |\tilde{b}_m^2|^2, \dot{a}_m^1 \dot{a}_m^2 \det(\tilde{b}_m \tilde{b}_m^\top)) \\
&\quad + o_p(b_n^{1/2} \ell_n^{-1}) \\
&= T b_n^{1/2} \ell_n^{-1} \left(\frac{1}{\sqrt{2}} \varphi(\dot{a}_m^1 |\tilde{b}_m^1|^2 + \dot{a}_m^2 |\tilde{b}_m^2|^2, \dot{a}_m^1 \dot{a}_m^2 \det(\tilde{b}_m \tilde{b}_m^\top)) - \sqrt{\dot{a}_m^1 |\tilde{b}_m^1|^2} - \sqrt{\dot{a}_m^2 |\tilde{b}_m^2|^2} \right) + o_p(b_n^{1/2} \ell_n^{-1}).
\end{aligned} \tag{5.16}$$

Moreover, Lemmas A.3 and 5.1 yield

$$\begin{aligned}
& \partial_\sigma^k \log \det(\tilde{D}_{j,m} \tilde{D}_{j,m,*}^{-1}) \\
&= \partial_\sigma^k \log \det(\mathcal{E} + \tilde{D}_{j,m,*}^{-1/2} (\tilde{D}_{j,m} - \tilde{D}_{j,m,*}) \tilde{D}_{j,m,*}^{-1/2}) = \sum_{p=1}^{\infty} \frac{(-1)^{p-1}}{p} \partial_\sigma^k \text{tr}((\tilde{D}_{j,m,*}^{-1} (\tilde{D}_{j,m} - \tilde{D}_{j,m,*}))^p) \\
&= \sum_{p=1}^{\infty} \frac{(-1)^{p-1}}{p} \partial_\sigma^k \text{tr}((\dot{D}_{j,m,*}^{-1} (\dot{D}_{j,m} - \dot{D}_{j,m,*}))^p) + o_p(b_n^{1/2} \ell_n^{-1}) = \partial_\sigma^k \log \det(\dot{D}_{j,m} \dot{D}_{j,m,*}^{-1}) + o_p(b_n^{1/2} \ell_n^{-1})
\end{aligned} \tag{5.17}$$

when $|\tilde{b}_{m,*}^j| \geq |\tilde{b}_m^j|$, where $\tilde{D}_{j,m,*}$ and $\dot{D}_{j,m,*}$ are obtained by substituting $\sigma = \sigma_*$ in $\tilde{D}_{j,m}$ and $\dot{D}_{j,m}$, respectively. Similarly, we have $\partial_\sigma^k \log \det(\tilde{D}_{j,m} \tilde{D}_{j,m,*}^{-1}) = \partial_\sigma^k \log \det(\dot{D}_{j,m} \dot{D}_{j,m,*}^{-1}) + o_p(b_n^{1/2} \ell_n^{-1})$ when $|\tilde{b}_{m,*}^j| < |\tilde{b}_m^j|$.

On the other hand, results in Section 4.2 yield

$$\begin{aligned}
\partial_\sigma^k \log \frac{\det \dot{D}_{j,m}}{\det \dot{D}_{j,m,*}} &= \frac{k_m^j}{\pi} \partial_\sigma^k \int_0^\pi \log \frac{c_j + 2(1 - \cos x)}{c_{j,*} + 2(1 - \cos x)} dx + o_p(b_n^{1/2} \ell_n^{-1}) \\
&= 2k_m^j \partial_\sigma^k \log \frac{\sqrt{c_j} + \sqrt{4 + c_j}}{\sqrt{c_{j,*}} + \sqrt{4 + c_{j,*}}} + o_p(b_n^{1/2} \ell_n^{-1}) \\
&= k_m^j \partial_\sigma^k (\sqrt{c_j} - \sqrt{c_{j,*}}) + o_p(b_n^{1/2} \ell_n^{-1}) = T b_n^{1/2} \ell_n^{-1} \sqrt{\dot{a}_m^j} \partial_\sigma^k (|b_m^j| - |b_{m,*}^j|) + o_p(b_n^{1/2} \ell_n^{-1}).
\end{aligned} \tag{5.18}$$

The residuals are bounded uniformly with respect to σ and m . Then we obtain $\sup_\sigma |b_n^{-1/2} \partial_\sigma^k (H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n)) - \partial_\sigma^k \mathcal{Y}_1(\sigma)| \rightarrow_p 0$ as $n \rightarrow \infty$ for any $\sigma \in \Lambda$ and $0 \leq k \leq 3$ by (5.1) and (5.13)–(5.18).

Finally, we obtain the results without [A1'] by using the arguments in Proposition 3.1 of Gloter and Jacod [13]. \square

6 Identifiability of the model

In this section, we check the identifiability condition, $\inf_{\sigma \neq \sigma_*} ((-\mathcal{Y}_1(\sigma))/|\sigma - \sigma_*|^2) > 0$ almost surely. This condition is necessary to deduce consistency of the maximum-likelihood-type estimator, as seen in Proposition 7.1. In general, it is not easy to check this condition directly because $\mathcal{Y}_1(\sigma)$ is a complicated function of b_t and a_t^j . On the other hand, Ogihara and Yoshida [26] proved that the identifiability condition [A3] of a model for equidistant observations without noise is sufficient for the identifiability of a model for nonsynchronous observations. This is also the case for our model.

Proposition 6.1. *Assume [A1], [A2], and [V]. Then there exists a positive constant c such that*

$$-\mathcal{Y}_1(\sigma) \geq \chi \int_0^T \{(|b_t^1|^2 - |b_{t,*}^1|^2)^2 + (|b_t^2|^2 - |b_{t,*}^2|^2)^2 + (b_t^1 \cdot b_t^2 - b_{t,*}^1 \cdot b_{t,*}^2)^2\} dt \tag{6.1}$$

for any σ , where

$$\chi = c(1 - \bar{\rho}^2) \frac{(v_{1,*} \wedge v_{2,*})^{1/2}}{v_{1,*} \vee v_{2,*}} \left(\sup_{j,t} a_t^j \right)^{-1/2} \left(\sup_{j,t,\sigma} (|b^j(t, X_t, \sigma)| \vee |b^j(t, X_t, \sigma)|^{-1}) \right)^{-5}.$$

In particular, $\inf_{\sigma \neq \sigma_*} ((-\mathcal{Y}_1(\sigma))/|\sigma - \sigma_*|^2) > 0$ almost surely under [A1]–[A3] and [V].

Proof. It is sufficient to show the results under the additional condition [A1'] by localization techniques similar to the proof of Proposition 2.1.

Let $\hat{D}_m = \tilde{D}_m - M_{m,*}$ and $\mathbf{B} = \sup_{j,t,\sigma} (|b^j(t, X_t, \sigma)| \vee |b^j(t, X_t, \sigma)|^{-1})$, then since

$$\begin{aligned} & u^\top \hat{D}_m^{-1/2} \tilde{S}_m \hat{D}_m^{-1/2} u \\ & \geq u^\top \left(\begin{array}{cc} |\tilde{b}_m^1|^2 \mathcal{E} & \tilde{b}_m^1 \cdot \tilde{b}_m^2 \{ |I_{i,m}^1 \cap I_{j,m}^2| |I_{i,m}^1|^{-1/2} |I_{j,m}^2|^{-1/2} \}_{i,j} \\ \tilde{b}_m^1 \cdot \tilde{b}_m^2 \{ |I_{i,m}^1 \cap I_{j,m}^2| |I_{i,m}^1|^{-1/2} |I_{j,m}^2|^{-1/2} \}_{j,i} & |\tilde{b}_m^2|^2 \mathcal{E} \end{array} \right) u \end{aligned}$$

for any $u \in \mathbb{R}^{\mathbf{J}_{1,n} + \mathbf{J}_{2,n}}$, we have $\|(\hat{D}_m^{1/2} \tilde{S}_m^{-1} \hat{D}_m^{1/2})^{1/2}\| \leq C\mathbf{B}(1 - \bar{\rho}^2)^{-1/2}$ by Lemma A.4, and hence we obtain

$$\begin{aligned} & \|(\hat{D}_m^{1/2} \tilde{S}_m^{-1} \hat{D}_m^{1/2})^{1/2} (\hat{D}_m^{-1/2} \tilde{S}_{m,*} \hat{D}_m^{-1/2}) (\hat{D}_m^{1/2} \tilde{S}_m^{-1} \hat{D}_m^{1/2})^{1/2}\| \\ & = \|\mathcal{E} + (\hat{D}_m^{1/2} \tilde{S}_m^{-1} \hat{D}_m^{1/2})^{1/2} (\hat{D}_m^{-1/2} (\tilde{S}_{m,*} - \tilde{S}_m) \hat{D}_m^{-1/2}) (\hat{D}_m^{1/2} \tilde{S}_m^{-1} \hat{D}_m^{1/2})^{1/2}\| \leq 1 + C\mathbf{B}^2(1 - \bar{\rho}^2)^{-1}. \end{aligned}$$

Then Lemma A.6 yields

$$\text{tr}(\tilde{S}_m^{-1} \tilde{S}_{m,*} - \mathcal{E}) + \log \det \tilde{S}_m - \log \det \tilde{S}_{m,*} \geq C\mathbf{B}^{-2}(1 - \bar{\rho}^2) \text{tr}(\tilde{S}_m^{-1} (\tilde{S}_{m,*} - \tilde{S}_m) \tilde{S}_m^{-1} (\tilde{S}_{m,*} - \tilde{S}_m)). \quad (6.2)$$

Therefore, we have

$$\begin{aligned} & \text{tr}(\tilde{S}_m^{-1} \tilde{S}_{m,*} - \mathcal{E}) + \log \det \tilde{S}_m - \log \det \tilde{S}_{m,*} \\ & \geq C\mathbf{B}^{-2}(1 - \bar{\rho}^2) \text{tr}(\tilde{D}_m^{-1} (\tilde{S}_{m,*} - \tilde{S}_m) \tilde{S}_m^{-1} (\tilde{S}_{m,*} - \tilde{S}_m)) \geq C\mathbf{B}^{-2}(1 - \bar{\rho}^2) \text{tr}(\tilde{D}_m^{-1} (\tilde{S}_{m,*} - \tilde{S}_m) \tilde{D}_m^{-1} (\tilde{S}_{m,*} - \tilde{S}_m)) \\ & = C\mathbf{B}^{-2}(1 - \bar{\rho}^2) \left\{ \sum_{j=1}^2 (|\tilde{b}_{m,*}^j|^2 - |\tilde{b}_m^j|^2)^2 \text{tr}(\tilde{D}_{j,m}^{-1} \tilde{D}_{j,m}' \tilde{D}_{j,m}^{-1} \tilde{D}_{j,m}') + 2(\tilde{b}_{m,*}^1 \cdot \tilde{b}_{m,*}^2 - \tilde{b}_m^1 \cdot \tilde{b}_m^2)^2 \text{tr}(\tilde{D}_{1,m}^{-1} \tilde{G} \tilde{D}_{2,m}^{-1} \tilde{G}^\top) \right\}. \end{aligned}$$

Hence it is sufficient to show that limsup of three quantities $\text{tr}(\tilde{D}_{j,m}^{-1} \tilde{D}_{j,m}' \tilde{D}_{j,m}^{-1} \tilde{D}_{j,m}')$ for $j = 1, 2$ and $\text{tr}(\tilde{D}_{1,m}^{-1} \tilde{G} \tilde{D}_{2,m}^{-1} \tilde{G}^\top)$ are estimated from below by positive random variables.

By Lemma 5.1 and (5.12) with a sampling scheme $S^{n,1} \equiv S^{n,2}$, we obtain

$$\begin{aligned} b_n^{-1/2} \text{tr}(\tilde{D}_{j,m}^{-1} \tilde{D}_{j,m}' \tilde{D}_{j,m}^{-1} \tilde{D}_{j,m}') & = b_n^{-1/2} \text{tr}(\dot{D}_{j,m}^{-1} \dot{D}_{j,m}' \dot{D}_{j,m}^{-1} \dot{D}_{j,m}') + \bar{R}_n(\ell_n^{-1}) = b_n^{-5/2} (\hat{a}_m^j)^{-2} \text{tr}(\dot{D}_{j,m}^{-2}) + \bar{R}_n(\ell_n^{-1}) \\ & = \frac{b_n^{-5/2}}{(\hat{a}_m^j)^2 v_{j,*}^2} I_2 \left(\frac{b_n^{-1} |\tilde{b}_m^j|^2}{\hat{a}_m^j v_{j,*}} \right) + \bar{R}_n(\ell_n^{-1}) = \frac{\pi \ell_n^{-1}}{4(\hat{a}_m^j)^{1/2} v_{j,*}^{1/2} |\tilde{b}_m^j|^3} + \bar{R}_n(\ell_n^{-1}). \end{aligned}$$

Moreover, Lemma 5.1 and (5.12) yield

$$\begin{aligned} b_n^{-1/2} \text{tr}(\tilde{D}_{1,m}^{-1} \tilde{G} \tilde{D}_{2,m}^{-1} \tilde{G}^\top) & = b_n^{-5/2} \frac{\hat{a}_m^2}{(\hat{a}_m^1)^3} \text{tr}(\dot{D}_{1,m}^{-1} \dot{D}_{2,m}^{-1}) + \bar{R}_n(\ell_n^{-1}) \\ & \geq \frac{b_n^{-5/2} \hat{a}_m^2}{(\hat{a}_m^1)^3 v_{1,*} v_{2,*}} \text{tr} \left(\left(\left(\left(\frac{|\tilde{b}_m^2|^2 b_n^{-1} \hat{a}_m^2}{v_{2,*} (\hat{a}_m^1)^2} \right) \vee \frac{|\tilde{b}_m^1|^2 b_n^{-1}}{\hat{a}_m^1 v_{1,*}} \right) \mathcal{E} + M_{1,m} \right)^{-2} \right) + \bar{R}_n(\ell_n^{-1}) \\ & = \frac{\ell_n^{-1} \hat{a}_m^2}{v_{1,*} v_{2,*} (\hat{a}_m^1)^3} \frac{\pi}{4} \left(\left(\left(\frac{|\tilde{b}_m^2|^2 \hat{a}_m^2}{v_{2,*} (\hat{a}_m^1)^2} \right) \vee \frac{|\tilde{b}_m^1|^2}{\hat{a}_m^1 v_{1,*}} \right)^{-3/2} \right) + \bar{R}_n(\ell_n^{-1}). \end{aligned}$$

Similarly, we obtain

$$b_n^{-1/2} \text{tr}(\tilde{D}_{1,m}^{-1} \tilde{G} \tilde{D}_{2,m}^{-1} \tilde{G}^\top) \geq \frac{\ell_n^{-1} \hat{a}_m^1}{v_{1,*} v_{2,*} (\hat{a}_m^2)^3} \frac{\pi}{4} \left(\left(\left(\frac{|\tilde{b}_m^1|^2 \hat{a}_m^1}{v_{1,*} (\hat{a}_m^2)^2} \right) \vee \frac{|\tilde{b}_m^2|^2}{\hat{a}_m^2 v_{2,*}} \right)^{-3/2} \right) + \bar{R}_n(\ell_n^{-1}). \quad (6.3)$$

Therefore, we obtain (6.1).

In particular, by Lemma 6 and Remark 4 in [26], there exists a positive-valued random variable \mathcal{R} such that

$$-\mathcal{Y}_1(\sigma) \geq \chi \mathcal{R}(-\mathcal{Y}_0(\sigma))$$

for any σ . Therefore we have $\inf_{\sigma \neq \sigma_*} ((-\mathcal{Y}_1(\sigma))/|\sigma - \sigma_*|^2) > 0$ almost surely under [A1]–[A3] and [V]. \square

7 Asymptotic mixed normality of the estimator

In this section we prove the consistency and asymptotic mixed normality of $\hat{\sigma}_n$. To obtain asymptotic mixed normality, we prove stable convergence of the score function $b_n^{-1/4} \partial_\sigma H_n(\sigma_*, v_*)$ by means of the martingale limit theorem for a mixed normal limit in Jacod [19]. We also use the idea by Jacod et al. [20] to adapt the limit theorem to models containing observation noise.

Consistency is an immediate consequence of Proposition 2.1 and the identifiability condition.

Proposition 7.1. *Assume [A1]–[A3] and [V]. Then $\hat{\sigma}_n \rightarrow^p \sigma_*$ as $n \rightarrow \infty$.*

Proof. Let ϵ, δ be arbitrary positive constants. By Proposition 2.1, we have $\sup_\sigma |H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n) - \mathcal{Y}_1(\sigma)| \rightarrow^p 0$ as $n \rightarrow \infty$. Moreover, Proposition 6.1 ensures that there exists $\eta > 0$ such that $P[\inf_{\sigma \neq \sigma_*} ((-\mathcal{Y}_1(\sigma))/|\sigma - \sigma_*|^2) \leq \eta] < \epsilon$. Since $H_n(\hat{\sigma}_n, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n) \geq 0$ by the definition of $\hat{\sigma}_n$, we obtain

$$P[|\hat{\sigma}_n - \sigma_*| \geq \delta] < P[\mathcal{Y}_1(\hat{\sigma}_n) \leq -\eta\delta^2] + \epsilon \leq P[\sup_\sigma |H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n) - \mathcal{Y}_1(\sigma)| \geq \eta\delta^2] + \epsilon < 2\epsilon$$

for sufficiently large n . \square

Proposition 7.2. *Assume [A1], [A2], and [V]. Then $b_n^{-1/4} \partial_\sigma H_n(\sigma_*, \hat{v}_n) \rightarrow^{s-\mathcal{L}} \Gamma_1^{1/2} \mathcal{N}$ as $n \rightarrow \infty$.*

Proof. It is sufficient to prove the results assuming the additional condition [A1'].

Since $b_n^{-1/4} \partial_\sigma \tilde{H}_n(\sigma_*, v_*) = -2^{-1} b_n^{-1/4} \sum_m \bar{E}_m[\tilde{Z}_m^\top \partial_\sigma \tilde{S}_m^{-1} \tilde{Z}_m] + o_p(1)$, we only need to check assumptions of Theorem 3.2 in Jacod [19] for $\mathcal{X}_m^n = -2^{-1} b_n^{-1/4} \bar{E}_m[\tilde{Z}_m^\top \partial_\sigma \tilde{S}_m^{-1} \tilde{Z}_m]$. For any $\epsilon > 0$, Lemma 4.3 yields

$$\sum_{m=1}^{\lfloor \ell_n t \rfloor} E_m[|\mathcal{X}_m^n|^2 1_{\{|\mathcal{X}_m^n| > \epsilon\}}] \leq \frac{C b_n^{-1}}{\epsilon^2} \sum_{m=1}^{\lfloor \ell_n t \rfloor} E_m[(\tilde{Z}_m^\top \partial_\sigma \tilde{S}_m^{-1} \tilde{Z}_m)^4] \rightarrow^p 0.$$

Moreover, it is easy to see that $\sum_{m=1}^{\lfloor \ell_n t \rfloor} E_m[\mathcal{X}_m^n (W_{s_m} - W_{s_{m-1}})] \rightarrow^p 0$.

Let N be a bounded martingale orthogonal to W_t . We will show $\sum_{m=1}^{\lfloor \ell_n t \rfloor} E_m[\mathcal{X}_m^n (N_{s_m} - N_{s_{m-1}})] \rightarrow^p 0$. Let \mathbf{N} be the set of finite sums of random variables $f(\mathbf{X}_T) \prod_{j=1}^l g_j(\epsilon_{i_j}^{n_j, k_j})$ where f and g_j are bounded Borel functions, \mathbf{X}_T is an $\mathcal{F}_T^{(0)}$ -measurable random variable, $n_1, \dots, n_l \in \mathbb{N}$, $1 \leq k_1, \dots, k_l \leq 2$, and $i_1, \dots, i_l \in \mathbb{Z}_+$. Since \mathbf{N} is dense in $L^1(\Omega, \mathcal{F}_T, P)$, Jacod [18] (4.15) ensures that the set \mathbf{N}' of linear combinations of martingales $\{E[N|\mathcal{F}_t]\}_{0 \leq t \leq T}$ with $N \in \mathbf{N}$ are dense in all bounded martingales orthogonal to W . Therefore, it is sufficient to show that

$$\sum_{m=1}^{\lfloor \ell_n t \rfloor} E_m[\mathcal{X}_m^n (N'_{s_m} - N'_{s_{m-1}})] \rightarrow^p 0 \quad (7.1)$$

for $N' \in \mathbf{N}'$.

Martingales in the form $N'_t = \int_0^t a_t^0 dW_t + \sum_k \int_0^t a_t^k dM_t^k$ with a bounded step function a_t^0 , bounded progressively measurable functions $\{a_t^k\}_t$ and bounded $\mathcal{F}_t^{(0)}$ -martingales $\{M_t^k\}_t$ orthogonal to W obviously satisfy (7.1) and are dense in the set of all bounded $\mathcal{F}_t^{(0)}$ -martingales. Therefore, (7.1) holds for any bounded $\mathcal{F}^{(0)}$ -martingale N' .

Moreover, let $N \in \mathbf{N}$, $N'_t = E[N|\mathcal{F}_t]$, and $\mathbf{T} = \{\alpha; \tilde{I}_{\alpha,m} \cap \{S_{i_{l'}}^{n_{l'},k_{l'}}\}_{l'=1}^l \neq \emptyset\}$, then we have

$$N'_t = E[f(\mathbf{X}_T)E[\prod_{j=1}^l g_j(\epsilon_{i_j}^{n_j,k_j})|\mathcal{F}_T^{(0)} \otimes \mathcal{F}_t^{(1)}]|\mathcal{F}_t] = E[\prod_{j=1}^l g_j(\epsilon_{i_j}^{n_j,k_j})|\mathcal{F}_{\inf_{\tilde{I}_{\alpha,m}}^{(1)}}^{(1)}]E[f(\mathbf{X}_T)|\mathcal{F}_t^{(0)}]$$

for $\alpha \notin \mathbf{T}$ and $t \notin \cup_{\alpha'} \tilde{I}_{\alpha',m}$. Therefore, we obtain

$$\begin{aligned} & |E_m[\mathcal{X}_m^n(N'_{s_m} - N'_{s_{m-1}})]| \\ &= \frac{1}{2}|E_m[\sum_{\alpha,\beta} (\mathbf{A}_m^\top \partial_\sigma \tilde{S}_{m,*} \mathbf{A}_m)_{\alpha,\beta}^{-1} \bar{E}_m[(\tilde{\epsilon}_{\alpha,m} - \dot{\epsilon}_{\alpha,m})(\tilde{\epsilon}_{\beta,m} - \dot{\epsilon}_{\beta,m})](N'_{s_m} - N'_{s_{m-1}})]| \\ &\leq \frac{1}{2}|E_m[\sum_{\alpha,\beta \in \mathbf{T}} (\mathbf{A}_m^\top \partial_\sigma \tilde{S}_{m,*} \mathbf{A}_m)_{\alpha,\beta}^{-1} \bar{E}_m[\tilde{\epsilon}_{\alpha,m} \tilde{\epsilon}_{\beta,m} - \dot{\epsilon}_{\alpha,m} \tilde{\epsilon}_{\beta,m} - \dot{\epsilon}_{\beta,m} \tilde{\epsilon}_{\alpha,m}](N'_{s_m} - N'_{s_{m-1}})]| \\ &\quad + \frac{1}{2}|E_m[\sum_{\alpha,\beta \notin \mathbf{T}} (\mathbf{A}_m^\top \partial_\sigma \tilde{S}_{m,*} \mathbf{A}_m)_{\alpha,\beta}^{-1} \bar{E}_m[\dot{\epsilon}_{\alpha,m} \dot{\epsilon}_{\beta,m}](N'_{s_m} - N'_{s_{m-1}})]| \rightarrow^p 0. \end{aligned}$$

Lemma 4.3 yields

$$E_m[(\mathcal{X}_m^n)^2] = \frac{b_n^{-1/2}}{4} \{E_m[(\tilde{Z}_m \partial_\sigma \tilde{S}_{m,*}^{-1} \tilde{Z}_m)^2] - E_m[\tilde{Z}_m \partial_\sigma \tilde{S}_{m,*}^{-1} \tilde{Z}_m]^2\} = \frac{b_n^{-1/2}}{2} \text{tr}(\tilde{S}_{m,*} \partial_\sigma \tilde{S}_{m,*}^{-1} \tilde{S}_{m,*} \partial_\sigma \tilde{S}_{m,*}^{-1}) + \bar{R}_n(b_n^{-1/2}).$$

On the other hand, since $\partial_\sigma \log \det \tilde{S}_m(x, \sigma) = -\text{tr}(\partial_\sigma \tilde{S}_m \tilde{S}_m^{-1})$, we have

$$E_m[\tilde{Z}_m^\top \partial_\sigma^2 \tilde{S}_m^{-1} \tilde{Z}_m + \partial_\sigma^2 \log \det \tilde{S}_m]|_{\sigma=\sigma_*} = \text{tr}(\partial_\sigma^2 \tilde{S}_{m,*}^{-1} \tilde{S}_{m,*}) - \text{tr}(\partial_\sigma^2 \tilde{S}_{m,*}^{-1} \tilde{S}_{m,*}) + \text{tr}(\tilde{S}_{m,*}^{-1} \partial_\sigma \tilde{S}_{m,*} \tilde{S}_{m,*}^{-1} \partial_\sigma \tilde{S}_{m,*}).$$

Therefore we have

$$\sum_{m=1}^{[\ell_n t]} E_m[(\mathcal{X}_m^n)^2] = -b_n^{-1/2} \sum_{m=1}^{[\ell_n t]} E_m[\tilde{Z}_m^\top \partial_\sigma^2 \tilde{S}_m^{-1} \tilde{Z}_m + \partial_\sigma^2 \log \det \tilde{S}_m]|_{\sigma=\sigma_*} \rightarrow^p -\partial_\sigma^2 \mathcal{Y}_1(\sigma_*, t),$$

where

$$\begin{aligned} \mathcal{Y}_1(\sigma, t) &= \int_0^t \left\{ \frac{\sum_{j=1}^2 (|b_s^j|^2 - |b_{s,*}^j|^2)(|b_s^{3-j}|^2 \sqrt{\tilde{a}_s^1 \tilde{a}_s^2} + \tilde{a}_s^j \sqrt{\det(b_s b_s^\top)}) - 2(b_s^1 \cdot b_s^2 - b_{s,*}^1 \cdot b_{s,*}^2) b_s^1 \cdot b_s^2 \sqrt{\tilde{a}_s^1 \tilde{a}_s^2}}{2\sqrt{2} \sqrt{\det(b_s b_s^\top)} \varphi(\tilde{a}_s^1 |b_s^1|^2 + \tilde{a}_s^2 |b_s^2|^2, \tilde{a}_s^1 \tilde{a}_s^2 \det(b_s b_s^\top))} \right. \\ &\quad \left. - \frac{\varphi(\tilde{a}_s^1 |b_s^1|^2 + \tilde{a}_s^2 |b_s^2|^2, \tilde{a}_s^1 \tilde{a}_s^2 \det(b_s b_s^\top)) - \varphi(\tilde{a}_s^1 |b_{s,*}^1|^2 + \tilde{a}_s^2 |b_{s,*}^2|^2, \tilde{a}_s^1 \tilde{a}_s^2 \det(b_{s,*} b_{s,*}^\top))}{2\sqrt{2}} \right\} ds. \end{aligned}$$

Then Theorem 2.1 in Jacod [19] yields $b_n^{-1/4} \partial_\sigma H_n(\sigma_*, \hat{v}_n) \xrightarrow{s-\mathcal{L}} \Gamma_1^{1/2} \mathcal{N}$. \square

Proof of Theorem 2.1. Since the parameter space Λ is open, there exists $\epsilon > 0$ such that $O(\epsilon, \sigma_*) = \{\sigma; |\sigma - \sigma_*| < \epsilon\} \subset \Lambda$. Then we have

$$-\partial_\sigma H_n(\sigma_*, \hat{v}_n) = \int_0^1 \partial_\sigma^2 H_n(\sigma_*, \hat{v}_n)(\sigma_* + t(\hat{\sigma}_n - \sigma_*))(\hat{\sigma}_n - \sigma_*) dt$$

for $\hat{\sigma}_n \in \Lambda$, by $\partial_\sigma H_n(\hat{\sigma}_n, \hat{v}_n) = 0$.

Hence we obtain $b_n^{1/4}(\hat{\sigma}_n - \sigma_*) = \tilde{\Gamma}_{1,n}^{-1} b_n^{-1/4} \partial_\sigma H_n(\sigma_*, \hat{v}_n)$ on $\{\det \tilde{\Gamma}_{1,n} \neq 0 \text{ and } \hat{\sigma}_n \in O(\epsilon, \sigma_*)\}$, where $\tilde{\Gamma}_{1,n} = -b_n^{-1/2} \int_0^1 \partial_\sigma^2 H_n(\sigma_* + t(\hat{\sigma}_n - \sigma_*)) dt$. Then since Propositions 2.1 and 7.1 yield $P[\det \tilde{\Gamma}_{1,n} = 0] \rightarrow 0$, $P[\hat{\sigma}_n \in O(\epsilon, \sigma_*)^c] \rightarrow 0$ and $\tilde{\Gamma}_{1,n}^{-1} 1_{\{\det \tilde{\Gamma}_{1,n} \neq 0\}} \xrightarrow{p} \Gamma_1^{-1}$, we have $b_n^{1/4}(\hat{\sigma}_n - \sigma_*) \xrightarrow{s-\mathcal{L}} \Gamma_1^{-1/2} \mathcal{N}$ as $n \rightarrow \infty$ by Proposition 7.2.

Moreover, Proposition 2.1 and Theorem 7.1 ensure that $\hat{\Gamma}_{1,n} \xrightarrow{p} \Gamma_1$, which completes the proof. \square

8 Proof of the LAN property

To obtain the LAN property of our model, the arguments in the proof of Theorem 2.1 are essential. Indeed, by using Propositions 2.1 and 7.2, we obtain a LAMN-type property of the quasi-log-likelihood function H_n with respect to σ : $H_n(\sigma_* + b_n^{-1/4}u_1, v_*) - H_n(\sigma_*, v_*) - u_1^\top b_n^{-1/4}\partial_\sigma H_n(\sigma_*, v_*) - u_1^\top b_n^{-1/2}\partial_\sigma^2 H_n(\sigma_*, v_*)u_1/2 \rightarrow^p 0$ as $n \rightarrow \infty$ for any $u_1 \in \mathbb{R}^d$, and $(b_n^{-1/4}\partial_\sigma H_n(\sigma_*, v_*), -b_n^{-1/2}\partial_\sigma^2 H_n(\sigma_*, v_*)) \rightarrow^{s-\mathcal{L}} (\Gamma_1^{1/2}\mathcal{N}, \Gamma_1)$, where \mathcal{N} is a d -dimensional standard normal random variable independent of \mathcal{F} . On the other hand, under the assumptions of Theorem 2.2, the *true* log-likelihood ratio $\log(dP_{\sigma_* + b_n^{-1/4}u_1, v_* + b_n^{-1/2}u_2, n}/dP_{\sigma_*, v_*, n})$ for $u_1 \in \mathbb{R}^d$ and $u_2 \in \mathbb{R}^2$ is obtained as $-(Z_1^\top S_1^{-1}Z_1 + \log \det S_1)/2$ if we set $k_n = b_n$. We cannot apply the argument of Section 5 to this quantity because the estimate $\ell_n \rightarrow \infty$ is essential there. Therefore, we follow the approaches by Gloter and Jacod [12] to show the LAN property. We set a ‘subexperiment’ and a ‘superexperiment’, which are obtained by respectively removing and adding observations from the original experiment. The likelihood functions of these experiments have similar properties to H_n , and therefore we can prove the LAN properties for these experiments with the same limit distribution. We can prove that these results lead us to the LAN property of the original one.

Let $\mathcal{Z} = (\mathbb{R}^8)^\mathbb{N}$, $\pi_i(z) = (x_i^{k,j}, t_i^k, e_i^k)_{j,k=1,2}$ for $i \in \mathbb{Z}_+$ and $z = (x_{i'}^{k,j}, t_{i'}^k, e_{i'}^k)_{i' \in \mathbb{Z}_+, j,k=1,2} \in \mathcal{Z}$. Let $\mathcal{H} = \mathfrak{B}(\{\pi_i^{-1}(A); i \in \mathbb{Z}_+, A \in \mathcal{B}(\mathbb{R}^8)\})$, $P'_{\sigma'_*, v'_*}$ be the induced probability measure on $(\mathcal{Z}, \mathcal{H})$ by $((Y_{S_i^{n,k}}^j 1_{\{i \leq \mathbf{J}_{k,n}\}}, S_i^{n,k} 1_{\{i \leq \mathbf{J}_{k,n}\}}, \epsilon_i^{n,k} 1_{\{i \leq \mathbf{J}_{k,n}\}})_{i \in \mathbb{Z}_+, j,k=1,2})$ with a true value (σ'_*, v'_*) . We can ignore the event $\min_{j,m} k_m^j \leq 0$.

Let $\mathcal{H}' = \mathfrak{B}(t_i^k; i \in \mathbb{Z}_+, k = 1, 2)$, $j_0^k = -1$, $j_m^k = \max\{i; t_i^k < s_m\} \vee 0$ ($1 \leq m \leq \ell_n$), $\mathbf{l}(0) = 1$, $\mathbf{l}(m) = \min\{k; t_i^k = \max_{i',k'}\{t_{i'}^{k'} < s_m\} \text{ for some } i\}$ for $1 \leq m \leq \ell_n$,

$$\begin{aligned}\mathcal{H}^{n,0} &= \mathfrak{B}((x_{i+1}^{k,k} + e_{i+1}^{k,k} - x_i^{k,k} - e_i^{k,k}) 1_{\{i \notin \{j_m^k\}_m\}}; i \in \mathbb{Z}_+, k = 1, 2) \bigvee \mathcal{H}', \\ \mathcal{H}^{n,1} &= \mathfrak{B}(x_i^{k,k} + e_i^{k,k}; i \in \mathbb{Z}_+, k = 1, 2) \bigvee \mathcal{H}', \\ \mathcal{H}^{n,2} &= \mathcal{H}^{n,1} \bigvee \mathfrak{B}(x_{j_m^{(m),j}}^{l(m),j}; 1 \leq m \leq \ell_n, j = 1, 2).\end{aligned}$$

Then we can see $\mathcal{H}^{n,0} \subset \mathcal{H}^{n,1} \subset \mathcal{H}^{n,2}$ and

$$\log(dP_{\sigma_u, v_u}/dP_{\sigma_*, v_*}) = \log(dP'_{\sigma_u, v_u}/dP'_{\sigma_*, v_*})|_{\mathcal{H}^{n,1}}. \quad (8.1)$$

Moreover, we obtain

$$\log \frac{dP'_{\sigma_u, v_u}}{dP'_{\sigma_*, v_*}} \Big|_{\mathcal{H}^{n,l}} ((Y_{S_i^{n,k}}^j 1_{\{i \leq \mathbf{J}_{k,n}\}}, S_i^{n,k} 1_{\{i \leq \mathbf{J}_{k,n}\}}, \epsilon_i^{n,k} 1_{\{i \leq \mathbf{J}_{k,n}\}})_{i \in \mathbb{Z}_+, j,k=1,2}) = H_n^{(l)}(\sigma_u, v_u) - H_n^{(l)}(\sigma_*, v_*) \quad (8.2)$$

for $l = 0$, where $Z_m^{(0)} = Z_m$ and $S_m^{(0)} = S_m$ for $2 \leq m \leq \ell_n$, $Z_1^{(0)}$ and $S_1^{(0)}$ are defined similarly, $H_n^{(0)}(\sigma, v) = -\sum_{m=1}^{\ell_n} \{ (Z_m^{(0)})^\top (S_m^{(0)})^{-1}(\sigma, v) Z_m^{(0)} + \log \det S_m^{(0)}(\sigma, v) \} / 2$, $\sigma_u = \sigma_* + b_n^{-1/4}u_1$ and $v_u = v_* + b_n^{-1/2}u_2$ for $u = (u_1, u_2) \in \mathbb{R}^d \times \mathbb{R}^2$. $\mathcal{H}^{n,0}$ and $\mathcal{H}^{n,2}$ are σ -fields for ‘subexperiment’ and ‘superexperiment’, respectively, while $\mathcal{H}^{n,1}$ is the one for the original one. Therefore, (8.2) means that our quasi-likelihood function H_n is equal to the log-likelihood function of ‘subexperiment’ except the term for $m = 1$.

To obtain similar formula to (8.2) for $l = 2$, let $\mathbf{R}_m = S_{K_m^1}^{n,1} \vee S_{K_m^2}^{n,2}$, $\tilde{\mathbf{Y}}_{m,-}^k = \tilde{Y}_{K_{m-1}^k+1}^k - Y_{\mathbf{R}_{m-1}}^k$,

$$\tilde{\mathbf{Y}}_{m,+}^k = \begin{cases} Y_{S_{K_m^k}^{n,k}}^k - \tilde{Y}_{K_m^k-1}^k & \text{if } S_{K_m^k}^{n,k} = \mathbf{R}_m \\ (\tilde{Y}^k(I_{k_m^k, m}^k), Y_{\mathbf{R}_m}^k - \tilde{Y}_{S_{K_m^k}^{n,k}}^k)^\top & \text{if } S_{K_m^k}^{n,k} < \mathbf{R}_m \end{cases}$$

$\mathbf{Y}_{m,0} = \epsilon_{K_m^k}^{n,k}$ if $S_{K_m^{3-k}}^{n,3-k} < \mathbf{R}_m$, $\mathbf{Y}_{m,0} = (\epsilon_{K_m^1}^{n,1}, \epsilon_{K_m^2}^{n,2})^\top$ if $S_{K_m^1}^{n,1} = S_{K_m^2}^{n,2}$, and

$$Z_m^{(2)} = (((\tilde{\mathbf{Y}}_{m,-}^k)^\top, (\tilde{Y}^k(I_{i,m}^k))_{1 \leq i < k_m^k}^\top, (\tilde{\mathbf{Y}}_{m,+}^k)^\top)_{k=1}^2, \mathbf{Y}_{m,0}^\top)^\top$$

for $2 \leq m \leq \ell_n$. Then Observations $((\tilde{Y}_i^k)_{k,i}, (S_i^{n,k})_{k,i}, (Y_{\mathbf{R}_m}^j)_{j,m})$ are equivalent to $Z_m^{(2)}$, and hence (8.2) holds for $l = 2$, where $\mathbf{E}(v) = v_{3-k}$, $k_m^{(2),k} = k_m^k + 1$, $k_m^{(2),3-k} = k_m^{3-k}$, and $I_{k_m^{(2),k}, m}^k = [S_{K_m^k}^{n,k}, \mathbf{R}_m)$ if $S_{K_m^k}^{n,k} < \mathbf{R}_m$, $\mathbf{E}(v) = \text{diag}(v_1, v_2)$ and $(k_m^{(2),1}, k_m^{(2),2}) = (k_m^1, k_m^2)$ if $S_{K_m^1}^{n,1} = S_{K_m^2}^{n,2}$,

$$(M_{j,m}^{(2)})_{ii'} = 2\delta_{ii'} - \delta_{|i-i'|=1} - \delta_{(i,i')=(1,1)} - \delta_{i=i'=k_m^{(2),j}},$$

$$S_m^{(2)}(\sigma, v) = \begin{pmatrix} \text{diag}((|b^1|^2|I_{i,m}^1|)_{1 \leq i \leq k_m^{(2),1}}) + v_1 M_{1,m}^{(2)} & \{b^1 \cdot b^2 |I_{i,m}^1 \cap I_{j,m}^2|\}_{1 \leq i \leq k_m^{(2),1}, 1 \leq j \leq k_m^{(2),2}} \\ \{b^1 \cdot b^2 |I_{i,m}^1 \cap I_{j,m}^2|\}_{1 \leq j \leq k_m^{(2),2}, 1 \leq i \leq k_m^{(2),1}} & \text{diag}((|b^2|^2|I_{j,m}^2|)_{1 \leq j \leq k_m^{(2),2}}) + v_2 M_{2,m}^{(2)} \end{pmatrix} \mathbf{E}(v)$$

for $2 \leq m \leq \ell_n$, $Z_1^{(2)}$, $M_1^{(2)}$, and $S_1^{(2)}$ are similarly defined, and $H_n^{(2)}(\sigma, v) = -\sum_{m=1}^{\ell_n} \{(Z_m^{(2)})^\top S_m^{(2)}(\sigma, v)^{-1} Z_m^{(2)} + \log \det S_m^{(2)}\}/2$.

The log-likelihood functions $H_n^{(0)}$ and $H_n^{(2)}$ of ‘subexperiment’ and ‘superexperiment’, respectively, have similar forms to that of H_n , and hence we can prove convergence of likelihood ratios. Gloter and Jacod [12] showed that convergence of likelihood ratios of ‘subexperiment’ and ‘superexperiment’ imply convergence of that of the original experiment. Here, we use a slight extension of their result. The proof is straightforward. Let $\mathbf{U}_{\sigma,v}^{n,l} = dP'_{\sigma,v}/dP'_{\sigma_*,v_*}|_{\mathcal{H}^{n,l}}$, $K \in \mathbb{N}$, and $\{\sigma_n^k\}_{n \in \mathbb{N}, 1 \leq k \leq K} \subset \Lambda$ and $\{v_n^l\}_{n \in \mathbb{N}, 1 \leq l \leq K} \subset (0, \infty) \times (0, \infty)$ be arbitrary sequences.

Theorem 8.1. *Suppose that $(\mathbf{U}_{\sigma_n^1, v_n^1}^{n,l}, \dots, \mathbf{U}_{\sigma_n^K, v_n^K}^{n,l})$ converges in law under $P_{\sigma_*, v_*}^{n,l}$ to a limit $Y = (Y^1, \dots, Y^K)$ with $0 < Y^k < \infty$ a.s. and $E[Y^k] = 1$ for $l = 0, 2$ and $1 \leq k \leq K$. Then the same convergence holds for $l = 1$.*

We first prove the LAN properties of ‘subexperiment’ and ‘superexperiment’. Then Theorem 8.1 leads to the LAN property of the original one. Taylor’s formula yields

$$\begin{aligned} & H_n^{(l)}(\sigma_u, v_u) - H_n^{(l)}(\sigma_*, v_*) \\ &= b_n^{-1/4} \partial_\sigma H_n^{(l)}(\sigma_*, v_*) \cdot u_1 + 2^{-1} b_n^{-1/2} u_1^\top \partial_\sigma^2 H_n^{(l)}(\sigma_*, v_*) u_1 + b_n^{-1/2} \partial_v H_n^{(l)}(\sigma_*, v_*) \cdot u_2 \\ & \quad + 2^{-1} b_n^{-1} u_2^\top \partial_v^2 H_n^{(l)}(\sigma_*, v_*) u_2 + \int_0^1 \int_0^1 \sum_{i,j} \partial_{v_i} \partial_{\sigma_j} H_n^{(l)}(\sigma_{tu}, v_{su}) b_n^{-3/4} u_{2,i} u_{1,j} ds dt \\ & \quad + \int_0^1 \frac{(1-t)^3}{2} \left(\sum_{i,j,k} \partial_{\sigma_i} \partial_{\sigma_j} \partial_{\sigma_k} H_n^{(l)}(\sigma_{tu}, v_*) u_{1,i} u_{1,j} u_{1,k} b_n^{-3/4} + \sum_{i,j,k} \partial_{v_i} \partial_{v_j} \partial_{v_k} H_n^{(l)}(\sigma_*, v_{tu}) u_{2,i} u_{2,j} u_{2,k} b_n^{-3/2} \right) dt. \end{aligned}$$

We examine the limit of each term on the right-hand side.

Lemma 8.1. *Assume [A1''], [A2], and [V]. Then*

1. $\sup_\sigma |b_n^{-1/2} \partial_\sigma^k (H_n^{(l)}(\sigma, v_*) - H_n^{(l)}(\sigma_*, v_*)) - \partial_\sigma^k \mathcal{Y}_1(\sigma)| \rightarrow^p 0$,
2. $\sup_v |b_n^{-1} \partial_v^k (H_n^{(l)}(\sigma_*, v) - H_n^{(l)}(\sigma_*, v_*)) - \partial_v^k \mathcal{Y}_2(v)| \rightarrow^p 0$,
3. $\sup_{\sigma,v} |b_n^{-3/4} \partial_\sigma \partial_v H_n^{(l)}(\sigma, v)| \rightarrow^p 0$

as $n \rightarrow \infty$ for $0 \leq k \leq 3$ and $l = 0, 2$.

Proof. 1. We obtain the results by a similar argument to the proof of Proposition 2.1 together with Lemma 4.1, the results in Section 8 of [12], and similar estimates to Lemmas 5.1 and 4.2. For any $\epsilon > 0$, $(\epsilon \mathcal{E} + M_{j,m}^{(2)})^{-1}$ has a similar decomposition to (4.4) by replacing p_{i-1}, \dots, p_j by p'_{i-1}, \dots, p'_j . Therefore, estimate for the quantity corresponding to Λ_1 is obtained since

$$((\epsilon \mathcal{E} + M_{j,m}^{(2)})^{-1})_{11} = \frac{\prod_{l=1}^{k_m^j-1} p'_l(\epsilon)}{(p'_{k_m^j}(\epsilon) - 1) \prod_{l=1}^{k_m^j-1} p'_l(\epsilon)} = O(b_n^{1/2}).$$

2. We first obtain

$$\begin{aligned}
& b_n^{-1} \partial_v^l H_n^{(l)}(\sigma, v) \\
&= -\frac{1}{2} b_n^{-1} \sum_m \left\{ E_m[(Z_m^{(l)})^\top \partial_v^l (S_m^{(l)})^{-1} Z_m^{(l)}] + \partial_v^l \log \det S_m^{(l)} \right\} - \frac{1}{2} b_n^{-1} \sum_m \bar{E}_m[(Z_m^{(l)})^\top \partial_v^l (S_m^{(l)})^{-1} Z_m^{(l)}] \\
&= -\frac{1}{2} b_n^{-1} \sum_m \left\{ \text{tr}(\partial_v^l (S_m^{(l)})^{-1} S_{m,*}^{(l)}) + \partial_v^l \log \det S_m^{(l)} \right\} + O_p \left(\left(b_n^{-2} \sum_m \text{tr}(\partial_v^l (S_m^{(l)})^{-1} S_{m,*}^{(l)}) \partial_v^l (S_m^{(l)})^{-1} S_{m,*}^{(l)} \right)^{1/2} \right) \\
&= -\frac{1}{2} b_n^{-1} \sum_m \left\{ \text{tr}(\partial_v^l (S_m^{(l)})^{-1} S_{m,*}^{(l)}) + \partial_v^l \log \det S_m^{(l)} \right\} + o_p(1).
\end{aligned}$$

Let $\tilde{D}_m^{(l)} = (\tilde{D}_{1,m}^{(l)}, \tilde{D}_{2,m}^{(l)})$ for $l = 0, 2$, $k_m^{(0),j} = k_m^j$ for $j = 1, 2$, $\tilde{D}_{1,m}^{(l)} = ((S_m^{(l)})_{i,i'})_{1 \leq i, i' \leq k_m^{(l),1}}$,

$$\tilde{D}_{2,m}^{(0)} = ((S_m^{(0)})_{j,j'})_{k_m^{(0),1} < j, j' \leq k_m^{(0),1} + k_m^{(0),2}}, \quad \tilde{D}_{2,m}^{(2)} = \text{diag}(((S_m^{(2)})_{j,j'})_{k_m^{(2),1} < j, j' \leq k_m^{(2),1} + k_m^{(2),2}}, \mathbf{E}(v)),$$

$$\hat{G}^{(l)} = (\tilde{D}_{1,m}^{(l)})^{-1/2} \{ |I_{i,m}^1 \cap I_{j,m}^2| 1_{\{j \leq k_m^{(l),2}\}} \}_{1 \leq i \leq k_m^{(l),1}, 1 \leq j \leq k_m^{(l),2}} (\tilde{D}_{2,m}^{(l)})^{-1/2},$$

where $\tilde{k}_m^{(0),2} = k_m^{(0),2}$ and $\tilde{k}_m^{(2),2}$ is the size of $\tilde{D}_{2,m}^{(2)}$. Then we obtain

$$\begin{aligned}
\text{tr}((S_m^{(l)})^{-1} S_{m,*}^{(l)}) &= \text{tr} \left((\tilde{D}_m^{(l)})^{-1/2} \begin{pmatrix} \mathcal{E} & \hat{G}^{(l)} \\ (\hat{G}^{(l)})^\top & \mathcal{E} \end{pmatrix}^{-1} (\tilde{D}_m^{(l)})^{-1/2} (\tilde{D}_{m,*}^{(l)})^{1/2} \begin{pmatrix} \mathcal{E} & \hat{G}_*^{(l)} \\ (\hat{G}_*^{(l)})^\top & \mathcal{E} \end{pmatrix} (\tilde{D}_{m,*}^{(l)})^{1/2} \right) \\
&= \sum_{p=0}^{\infty} \left\{ \text{tr}((\tilde{D}_{1,m}^{(l)})^{-1/2} (\hat{G}^{(l)} (\hat{G}^{(l)})^\top)^p (\tilde{D}_{1,m}^{(l)})^{-1/2} \tilde{D}_{1,m,*}^{(l)} \right. \\
&\quad \left. - (\tilde{D}_{1,m}^{(l)})^{-1/2} (\hat{G}^{(l)} (\hat{G}^{(l)})^\top)^p \hat{G}^{(l)} (\tilde{D}_{2,m}^{(l)})^{-1/2} (\tilde{D}_{2,m,*}^{(l)})^{1/2} (\hat{G}_*^{(l)})^\top (\tilde{D}_{1,m,*}^{(l)})^{1/2} \right. \\
&\quad \left. + \text{tr}((\tilde{D}_{2,m}^{(l)})^{-1/2} ((\hat{G}^{(l)})^\top \hat{G}^{(l)})^p (\tilde{D}_{2,m}^{(l)})^{-1/2} \tilde{D}_{2,m,*}^{(l)} \right. \\
&\quad \left. - (\tilde{D}_{2,m}^{(l)})^{-1/2} (\hat{G}^{(l)})^\top (\hat{G}^{(l)} (\hat{G}^{(l)})^\top)^p (\tilde{D}_{1,m}^{(l)})^{-1/2} ((\tilde{D}_{1,m,*}^{(l)})^{1/2} \hat{G}_*^{(l)} (\tilde{D}_{2,m,*}^{(l)})^{1/2}) \right\}.
\end{aligned}$$

Since $\|(\tilde{D}_{j,m}^{(l)})^{-1} \tilde{D}_{j,m,*}^{(l)}\| = O_p(1)$, terms involving \hat{G} are $O_p(b_n^{1/2} \ell_n^{-1})$. Therefore we have

$$\begin{aligned}
& b_n^{-1} \text{tr}((S_m^{(l)})^{-1} S_{m,*}^{(l)}) \\
&= b_n^{-1} \sum_{j=1}^2 \text{tr}((\tilde{D}_{j,m}^{(l)})^{-1} \tilde{D}_{j,m,*}^{(l)}) + o_p(\ell_n^{-1}) \\
&= b_n^{-1} \sum_{j=1}^2 \frac{v_{j,*}}{v_j} \text{tr}(\mathcal{E}_{k_m^j} - (\tilde{D}_{j,m}^{(l)})^{-1} (\tilde{D}_{j,m}^{(l)} - v_j v_{j,*}^{-1} \tilde{D}_{j,m,*}^{(l)})) + o_p(\ell_n^{-1}) = \ell_n^{-1} \sum_{j=1}^2 \hat{a}_m^j \frac{v_{j,*}}{v_j} + o_p(\ell_n^{-1}).
\end{aligned}$$

Similarly we have $b_n^{-1} \text{tr}(\partial_v^k (S_m^{(l)})^{-1} S_{m,*}^{(l)}) = \ell_n^{-1} \sum_{j=1}^2 \hat{a}_m^j \partial_v^k \frac{v_{j,*}}{v_j} + o_p(\ell_n^{-1})$. Moreover, we obtain

$$\begin{aligned}
& b_n^{-1} \partial_v^k \log \frac{\det S_m^{(l)}}{\det S_{m,*}^{(l)}} \\
&= \sum_{j=1}^2 b_n^{-1} \partial_v^k \log \det((\tilde{D}_{j,m}^{(l)})^{-1} \tilde{D}_{j,m}^{(l)}) + b_n^{-1} \partial_v^k \log \det(\mathcal{E} - \hat{G}^{(l)} (\hat{G}^{(l)})^\top) - b_n^{-1} \partial_v^k \log \det(\mathcal{E} - \hat{G}_*^{(l)} (\hat{G}_*^{(l)})^\top) \\
&= b_n^{-1} \sum_{j=1}^2 \partial_v^k \log \det(v_j v_{j,*}^{-1} \mathcal{E}_{k_m^j} + (\tilde{D}_{j,m}^{(l)})^{-1} (\tilde{D}_{j,m}^{(l)} - v_j v_{j,*}^{-1} \tilde{D}_{j,m,*}^{(l)})) + o_p(\ell_n^{-1}) \\
&= \ell_n^{-1} \sum_{j=1}^2 \hat{a}_m^j \partial_v^k \log(v_j v_{j,*}^{-1}) + o_p(\ell_n^{-1}).
\end{aligned}$$

3. Since $\partial_v \log \det S_m^{(i)} = -\text{tr}(\partial_v S_m^{(i)} (S_m^{(i)})^{-1})$ and

$$b_n^{-3/4} \partial_\sigma \partial_v H_n^{(i)}(\sigma, v) = -\frac{1}{2} b_n^{-3/4} \sum_m \{ \text{tr}(\partial_\sigma \partial_v (S_m^{(i)})^{-1} S_{m,*}^{(i)}) + \partial_\sigma \partial_v \log \det S_m^{(i)} \} + o_p(1),$$

we have $b_n^{-3/4} \partial_\sigma \partial_v H_n^{(i)}(\sigma, v) = o_p(1)$. Sobolev's inequality and similar estimates for $\partial_\sigma \partial_v^2$ and $\partial_\sigma^2 \partial_v$ yield the results. \square

The following lemma completes the proof of the LAN properties of 'subexperiment' and 'superexperiment'.

Lemma 8.2. *Assume $[A1'']$, $[A2]$, and $[V]$. Then $(b_n^{-1/4} \partial_\sigma \hat{H}_n^{(l)}(\sigma_*, v_*), b_n^{-1/2} \partial_v \hat{H}_n^{(l)}(\sigma_*, v_*)) \rightarrow^{s-\mathcal{L}} \text{diag}(\Gamma_1^{1/2}, \Gamma_2^{1/2}) \tilde{\mathcal{N}}$ for $l = 0, 2$, where $\tilde{\mathcal{N}}$ is a $(d+2)$ -dimensional normal random variable independent of \mathcal{F} .*

Proof. Let $\tilde{\mathcal{X}}_m^n = -b_n^{-1/4} \bar{E}_m[(Z_m^{(l)})^\top \partial_\sigma (S_{m,*}^{(l)})^{-1} Z_m^{(l)}]/2 - b_n^{-1/2} \bar{E}_m[(Z_m^{(l)})^\top \partial_v (S_{m,*}^{(l)})^{-1} Z_m^{(l)}]/2$, then we have

$$\begin{aligned} \sum_{m=1}^{[\ell_n t]} E_m[|\tilde{\mathcal{X}}_m^n|^2 1_{\{|\mathcal{X}_m^n| > \epsilon\}}] &\leq \frac{C}{\epsilon^2} b_n^{-1} \sum_m E_m[(Z_m^{(l)})^\top \partial_\sigma (S_{m,*}^{(l)})^{-1} Z_m^{(l)}]^4 + \frac{C}{\epsilon^2} b_n^{-2} \sum_m E_m[(Z_m^{(l)})^\top \partial_v (S_{m,*}^{(l)})^{-1} Z_m^{(l)}]^4 \\ &\leq \frac{C b_n^{-2}}{\epsilon^2} \sum_m \sum_{j=1}^4 \text{tr}((\partial_v (S_{m,*}^{(l)})^{-1} S_{m,*}^{(l)})^j) + o_p(1) \rightarrow^p 0. \end{aligned}$$

Moreover, similarly to the proof of Proposition 7.2, we have

$$\sum_{m=1}^{[\ell_n t]} E_m[\tilde{\mathcal{X}}_m^n (N_{s_m} - N_{s_{m-1}})] = \sum_{m=1}^{[\ell_n t]} E_m[\tilde{\mathcal{X}}_m^n (W_{s_m} - W_{s_{m-1}}, W'_{s_m} - W'_{s_{m-1}})] = 0$$

for any bounded martingale N orthogonal to $(W_t, W'_t)_t$.

Therefore, by Theorem 3.2 in Jacod [19], it is sufficient to show that

$$\sum_{m=1}^{[\ell_n t]} E_m[(\tilde{\mathcal{X}}_m^n)^2] \rightarrow^p \text{diag}(-\partial_\sigma^2 \mathcal{Y}_1(\sigma_*, t), -\partial_v^2 \mathcal{Y}_2(v_*, t)),$$

where $\mathcal{Y}_2(v, t) = -\int_0^t \sum_{j=1}^2 a_s^j \{(v_{j,*}/v_j) - 1 + \log(v_j/v_{j,*})\} ds/2$.

Then we obtain the desired results by

$$\begin{aligned} \sum_m E_m[(\tilde{\mathcal{X}}_m^n)^2] &= \frac{1}{4} b_n^{-1/2} \sum_m E_m[\bar{E}_m[(Z_m^{(l)})^\top (\partial_\sigma (S_{m,*}^{(l)})^{-1} + b_n^{-1/4} \partial_v (S_{m,*}^{(l)})^{-1}) Z_m^{(l)}]^2] \\ &= \frac{b_n^{-1/2}}{2} \sum_m \text{tr}(S_{m,*}^{(l)} (\partial_\sigma (S_{m,*}^{(l)})^{-1} + b_n^{-1/4} \partial_v (S_{m,*}^{(l)})^{-1}) S_{m,*}^{(l)} (\partial_\sigma (S_{m,*}^{(l)})^{-1} + b_n^{-1/4} \partial_v (S_{m,*}^{(l)})^{-1})) \\ &= -b_n^{-1/2} \partial_\sigma^2 H_n^{(l)}(\sigma_*, v_*) - b_n^{-1} \partial_v^2 H_n^{(l)}(\sigma_*, v_*) + \frac{b_n^{-3/4}}{2} \sum_m \text{tr}(\partial_\sigma S_{m,*}^{(l)} (S_{m,*}^{(l)})^{-1} \partial_v S_{m,*}^{(l)} (S_{m,*}^{(l)})^{-1}) \\ &\rightarrow^p \text{diag}(-\partial_\sigma^2 \mathcal{Y}_1(\sigma_*, t), -\partial_v^2 \mathcal{Y}_2(v_*, t)). \end{aligned}$$

\square

Proof of Theorem 2.2. Let $\mathbf{U}(u) = \exp(u^\top \Gamma^{1/2} \tilde{\mathcal{N}} - u^\top \Gamma u/2)$ for $u \in \mathbb{R}^{d+2}$. Let $Z^{(1)} = (\epsilon_0^{n,k}, (\tilde{Y}_i^k - \tilde{Y}_{i-1}^k)_{i=1}^{\mathbf{J}_{k,n}})_{k=1,2}$, $S^{(1)}(\sigma, v)$ be a symmetric matrix of size $\mathbf{J}_{1,n} + \mathbf{J}_{2,n} + 2$ defined by $(S^{(1)}(\sigma, v))_{11} = v_1$, $(S^{(1)}(\sigma, v))_{\mathbf{J}_{1,n}+2, \mathbf{J}_{1,n}+2} = v_2$,

$$\begin{aligned} (S^{(1)}(\sigma, v))_{ij} &= \text{diag}(v_1 M(\mathbf{J}_{1,n} + 1), v_2 M(\mathbf{J}_{2,n} + 1))_{ij} && \text{if } i \neq j \text{ and } \{i, j\} \cap \{1, \mathbf{J}_{1,n} + 2\} \neq \emptyset, \\ (S^{(1)}(\sigma, v))_{ij} &= |b^1(\sigma)|^2 (S_{i-1}^{n,1} - S_{i-2}^{n,1}) \delta_{ij} + v_1 M(\mathbf{J}_{1,n} + 1)_{ij} && \text{if } 2 \leq i, j \leq \mathbf{J}_{1,n} + 1, \\ (S^{(1)}(\sigma, v))_{ij} &= |b^2(\sigma)|^2 (S_{i'-1}^{n,2} - S_{i'-2}^{n,2}) \delta_{ij} + v_2 M(\mathbf{J}_{2,n} + 1)_{i'j'} && \text{if } 2 \leq i', j' \leq \mathbf{J}_{2,n} + 1, \\ (S^{(1)}(\sigma, v))_{ij} &= b^1 \cdot b^2(\sigma) (S_{i-1}^{n,1} \wedge S_{j'-1}^{n,2} - S_{i-2}^{n,1} \vee S_{j'-2}^{n,2})_+ && \text{if } 2 \leq i \leq \mathbf{J}_{1,n} + 1 \text{ and } 2 \leq j' \leq \mathbf{J}_{2,n} + 1, \end{aligned}$$

where $i' = i - \mathbf{J}_{1,n} - 1$ and $j' = j - \mathbf{J}_{1,n} - 1$. Then we have (8.2) for $l = 1$ with $H_n^{(1)}(\sigma, v) = -((Z^{(1)})^\top (S^{(1)}(\sigma, v))^{-1} Z^{(1)} + \log \det S^{(1)}(\sigma, v))/2$. Moreover, Theorem 8.1 and Lemmas 8.1 and 8.2 yield

$$(H_n^{(1)}(\sigma_{u^{(1)}}, v_{u^{(1)}}) - H_n^{(1)}(\sigma_*, v_*), \dots, H_n^{(1)}(\sigma_{u^{(k)}}, v_{u^{(k)}}) - H_n^{(1)}(\sigma_*, v_*)) \rightarrow^d (\log \mathbf{U}(u^{(1)}), \dots, \log \mathbf{U}(u^{(k)})) \quad (8.3)$$

as $n \rightarrow \infty$ for $u^{(1)}, \dots, u^{(k)} \in \mathbb{R}^{d+2}$.

Furthermore, similar estimates to the proof of Lemma 8.1 yield $\sup_{\sigma, v} |b_n^{-3/4} \partial_\sigma \partial_v H_n^{(1)}(\sigma, v)| \rightarrow^p 0$, $\sup_\sigma |b_n^{-3/4} \partial_\sigma^3 H_n^{(1)}(\sigma, v_*)| \rightarrow^p 0$, and $\sup_v |b_n^{-3/2} \partial_v^3 H_n^{(1)}(\sigma_*, v)| \rightarrow^p 0$. Therefore we obtain

$$H_n^{(1)}(\sigma_u, v_u) - H_n^{(1)}(\sigma_*, v_*) - (u \cdot \mathbf{V}_{1,n} - u^\top \mathbf{V}_{2,n} u / 2) \rightarrow^p 0 \quad (8.4)$$

as $n \rightarrow \infty$ for any $u \in \mathbb{R}^{d+2}$, where $\mathbf{V}_{1,n} = (b_n^{-1/4} \partial_\sigma H_n^{(1)}(\sigma_*, v_*), b_n^{-1/2} \partial_v H_n^{(1)}(\sigma_*, v_*))$ and $\mathbf{V}_{2,n} = -\text{diag}(b_n^{-1/2} \partial_\sigma^2 H_n^{(1)}(\sigma_*, v_*), b_n^{-1} \partial_v^2 H_n^{(1)}(\sigma_*, v_*))$. (8.3) and (8.4) yield $\mathbf{V}_{1,n} \rightarrow^d \Gamma^{1/2} \tilde{\mathcal{N}}$ and $\mathbf{V}_{2,n} \rightarrow^p \Gamma$, and therefore we obtain the LAN property of the original experiment with $\Gamma_n = \mathbf{V}_{2,n}$ and $\mathcal{N}_n = \Gamma^{-1/2} \mathbf{V}_{1,n}$ by (8.1). \square

9 Proof of the results in Section 2.4

In this final section, we complete the proof of remaining results in Section 2. Proposition 2.2 is proven by the scheme of Yoshida [30, 31]. Proposition 6.1 and moment estimates in Lemmas 4.4 and 5.2 enable us to check the assumptions of Theorem 2 in [31]. Then the results on convergence of moments and the Bayes-type estimator are obtained by Proposition 2.2.

Outline of the proof of Proposition 2.2. We apply Theorem 2 in Yoshida [31]. It is sufficient to prove the following five conditions for any $L > 0$ with some positive constant δ_1 and δ_2 :

1. There exists $C_L > 0$ such that $P[\inf_{\sigma \neq \sigma_*} (-\mathcal{Y}_1(\sigma)/|\sigma - \sigma_*|^2) \leq r^{-1}] \leq C_L/r^L$ and $P[\{r^{-1}|u|^2 \leq u^\top \Gamma_1 u/4 \text{ for any } u \in \mathbb{R}^d\}^c] \leq C_L/r^L$ for any $r > 0$.
2. $\sup_n E[(b_n^{-1/4} |\partial_\sigma H_n(\sigma_*, \hat{v}_n)|)^L] < \infty$.
3. $\sup_n E[(b_n^{\delta_1} \sup_\sigma |b_n^{-1/2} (H_n(\sigma, \hat{v}_n) - H_n(\sigma_*, \hat{v}_n)) - \mathcal{Y}_1(\sigma)|)^L] < \infty$.
4. $\sup_n E[(b_n^{-1/2} \sup_\sigma |\partial_\sigma^3 H_n(\sigma, \hat{v}_n)|)^L] < \infty$.
5. $\sup_n E[(b_n^{\delta_2} |b_n^{-1/2} \partial_\sigma^2 H_n(\sigma_*, \hat{v}_n) + \Gamma_1|)^L] < \infty$.

By Taylor's formula for $\mathcal{Y}_1(\sigma)$ and relations $\mathcal{Y}_1(\sigma_*) = \partial_\sigma \mathcal{Y}_1(\sigma_*) = 0$, we obtain $\inf_{\sigma \neq \sigma_*} (-\mathcal{Y}_1(\sigma)/|\sigma - \sigma_*|^2) \leq \inf_{u \in \mathbb{R}^d \setminus \{0\}} u^\top \Gamma_1 u / (2|u|^2)$. Then Proposition 6.1 and [B3] yield point 1. By Lemmas 4.4 and 5.2 and a similar argument to the proof of Proposition 2.1, we obtain 3–5 and $\sup_n E[(b_n^{-1/4} |\partial_\sigma H_n(\sigma_*, \hat{v}_n) - \partial_\sigma \tilde{H}_n(\sigma_*, v_*)|)^L] < \infty$. Moreover, by the Burkholder–Davis–Gundy inequality, we obtain

$$\begin{aligned} & E[|b_n^{-1/4} \partial_\sigma \tilde{H}_n(\sigma_*, v_*)|^L] \\ &= E\left[\left|\frac{b_n^{-1/4}}{2} \sum_m \bar{E}_m[\tilde{Z}_m^\top \partial_\sigma \tilde{S}_{m,*}^{-1} \tilde{Z}_m]\right|^L\right] \leq CE\left[\left(b_n^{-\frac{1}{2}} \sum_m \bar{E}_m[\tilde{Z}_m^\top \partial_\sigma \tilde{S}_{m,*}^{-1} \tilde{Z}_m]^2\right)^{L/2}\right] \\ &\leq CE\left[\left(b_n^{-\frac{1}{2}} \sum_m E_m[\bar{E}_m[\tilde{Z}_m^\top \partial_\sigma \tilde{S}_{m,*}^{-1} \tilde{Z}_m]^2]\right)^{L/2}\right] + CE\left[\left(b_n^{-1} \sum_m \bar{E}_m[\bar{E}_m[\tilde{Z}_m^\top \partial_\sigma \tilde{S}_{m,*}^{-1} \tilde{Z}_m]^2]^2\right)^{L/4}\right] \\ &\leq CE\left[\left(b_n^{-\frac{1}{2}} \sum_m \text{tr}((\partial_\sigma \tilde{S}_{m,*}^{-1} \tilde{S}_{m,*})^2)\right)^{L/2}\right] + CE\left[\left(b_n^{-1} \sum_m E_m[(\tilde{Z}_m^\top \partial_\sigma \tilde{S}_{m,*}^{-1} \tilde{Z}_m)^L]^{4/L}\right)^{L/4}\right] \\ &= O((b_n^{-1/2} \ell_n b_n^{1/2} \ell_n^{-1})^{L/2}) + E[\bar{R}_n((b_n^{-5} k_n^8 \ell_n)^{L/4})] = O(1), \end{aligned}$$

which implies point 2. \square

Proof of Theorem 2.3.

We extend $\mathbf{Z}_n(u)$ to a continuous function on \mathbb{R}^d satisfying $\lim_{|u| \rightarrow \infty} \mathbf{Z}(u) = 0$ with the supremum norm of the extended function the same as for the original one. Then by Theorem 5 and Remark 5 in Yoshida [31], it is sufficient to show $\limsup_{n \rightarrow \infty} E[|b_n^{1/4}(\hat{\sigma}_n - \sigma_*)|^p] < \infty$ for any $p > 0$ and $\mathbf{Z}_n \rightarrow^{s-\mathcal{L}} \mathbf{Z}$ in $C(B(R))$ as $n \rightarrow \infty$ for any $R > 0$, where $\mathbf{Z}(u) = \exp(\mathcal{N} \cdot u - u^\top \Gamma_1 u / 2)$ and $B(R) = \{u; |u| \leq R\}$.

By Lemma 4.4 and a similar argument to the proof of Proposition 2.1, we have

$$\sup_n E \left[\sup_{u \in C(B(R))} |\partial_u \log \mathbf{Z}_n(u)| \right] < \infty.$$

Then Propositions 2.1 and 7.2 and tightness criterion in C space in Billingsley [6] yield $\log \mathbf{Z}_n \rightarrow^{s-\mathcal{L}} \log \mathbf{Z}$ in $C(B(R))$. Then (2.10) completes the proof. \square

Proof of Theorem 2.4.

By Theorem 10 in Yoshida [31], it is sufficient to show

$$\sup_n E_n \left[\left(\int_{U_n} \mathbf{Z}_n(u) \pi(\sigma_* + b_n - 1/4u) du \right)^{-1} \right] < \infty. \quad (9.1)$$

By Proposition 2.1, we obtain $\sup_n E[|H_n(\sigma_* + b_n^{-1/4}u) - H_n(\sigma_*)|^p] \leq C_p |u|^p$ for any $U(\delta)$, where $U(\delta) = \{u \in \mathbb{R}^d; |u_i| \leq \delta (i = 1, \dots, d)\}$. Then we have (9.1) by Lemma 2 in [31]. \square

A Appendix

A.1 Results from linear algebra

Lemma A.1. *Let A and B be matrices, with A nonnegative definite and symmetric. Then*

$$|\text{tr}(AB)| \leq \text{tr}(A) \|B\|.$$

Lemma A.2. *Let $l \in \mathbb{N}$, A^j and B^j be real-valued matrices and $\{\lambda_k^j\}_k$ be eigenvalues of A^j for $1 \leq j \leq l$. Assume that A^j is symmetric and all the elements of B^j are nonnegative for $1 \leq j \leq l$. Then*

$$\sum_{i_1, \dots, i_{2l}} \prod_{j=1}^l \left(|A_{i_{2j-1}, i_{2j}}^j| B_{i_{2j}, i_{2j+1}}^j \right) \leq \prod_{j=1}^l \left(\|B^j\| \sum_k |\lambda_k^j| \right),$$

where $i_{2l+1} = i_1$.

Proof. Let U^j be an orthogonal matrix such that $A^j = (U^j)^\top \text{diag}((\lambda_k^j)_k) U^j$. Then

$$\begin{aligned} \sum_{i_1, \dots, i_{2l}} \prod_{j=1}^l \left(|A_{i_{2j-1}, i_{2j}}^j| B_{i_{2j}, i_{2j+1}}^j \right) &\leq \sum_{k_1, \dots, k_l} \sum_{i_1, \dots, i_{2l}} \prod_{j=1}^l \left(|\lambda_{k_j}^j| |U_{k_j, i_{2j-1}}^j| |U_{i_{2j}, k_j}^j| B_{i_{2j}, i_{2j+1}}^j \right) \\ &\leq \sum_{k_1, \dots, k_l} \prod_{j=1}^l \left\{ |\lambda_{k_j}^j| \|B^j\| \sum_i (U_{k_j, i}^j)^2 \right\} = \prod_{j=1}^l \left(\|B^j\| \sum_k |\lambda_k^j| \right). \end{aligned}$$

\square

Lemma A.3. *Let A be a symmetric matrix with $\|A\| < 1$. Then $\log \det(\mathcal{E} + A) = \sum_{p=1}^{\infty} (-1)^{p-1} p^{-1} \text{tr}(A^p)$.*

Proof. Let $\{\lambda_j\}_{j=1}^k$ be eigenvalues of A . Then $\sup_j |\lambda_j| = \|A\| < 1$, and hence

$$\log \det(\mathcal{E} + A) = \sum_j \log(1 + \lambda_j) = \sum_j \sum_{p=1}^{\infty} (-1)^{p-1} p^{-1} \lambda_j^p = \sum_{p=1}^{\infty} (-1)^{p-1} p^{-1} \text{tr}(A^p).$$

\square

Lemma A.4. Let A and B be symmetric, positive definite matrices. Assume that $v^\top A v \geq v^\top B v$ for any vector v . Then $\|A^{-1}\| \leq \|B^{-1}\|$ and $\|A^{-1/2}\| \leq \|B^{-1/2}\|$.

Proof. Let $(\lambda_j^A)_j$ and $(\lambda_j^B)_j$ be eigenvalues of A and B , respectively. Then for any unit vector v , there exists an orthogonal matrix U such that

$$\sum_j \lambda_j^A v_j^2 \geq \sum_j \lambda_j^B (Uv)_j^2 \geq \inf_j \lambda_j^B.$$

Therefore we obtain $\|A^{-1}\|^{-1} = \inf_j \lambda_j^A \geq \inf_j \lambda_j^B = \|B^{-1}\|^{-1}$ and $\|A^{-1/2}\|^{-1} = \inf_j (\lambda_j^A)^{1/2} \geq \inf_j (\lambda_j^B)^{1/2} = \|B^{-1/2}\|^{-1}$. \square

Lemma A.5. Let B be a symmetric, positive definite matrix and A be a symmetric, nonnegative definite matrix. Then $\text{tr}(AB) \geq \text{tr}(A)\|B^{-1}\|^{-1}$.

Proof. Let $\{\lambda_j^A\}_j$ and $\{\lambda_j^B\}_j$ be eigenvalues of A and B , respectively, and U be an orthogonal matrix satisfying $UAU^\top = \text{diag}((\lambda_j^A)_j)$. Then since $(UBU^\top)_{jj} \geq \inf_j \lambda_j^B = \|B^{-1}\|^{-1}$, we obtain

$$\text{tr}(AB) = \sum_j \lambda_j^A (UBU^\top)_{jj} \geq \sum_j \lambda_j^A \|B^{-1}\|^{-1} = \text{tr}(A)\|B^{-1}\|^{-1}.$$

\square

Lemma A.6. Let $\eta > 0$ and A be a symmetric matrix. Assume that $\mathcal{E} + A$ is positive definite and $\|\mathcal{E} + A\| \leq \eta$. Then $\text{tr}(A) - \log \det(\mathcal{E} + A) \geq \text{tr}(A^2)/(4\eta + 4)$.

Proof. We easily obtain the results by using the fact that $\log \det(\mathcal{E} + A) = \sum_k \log(1 + \lambda_k)$ and that $x - x^2/(4\eta + 4) \geq \log(1 + x)$ for $-1 < x \leq \eta + 1$, where $(\lambda_j)_j$ are eigenvalues of A . \square

Lemma A.7. Let A be a symmetric matrix, B a matrix of suitable size and $(\lambda_j)_j$ eigenvalues of $B^\top AB$. Then

1. $|(B^\top AB)_{ii}| \leq \|A\|(B^\top B)_{ii}$ for any i .
2. $\sum_j |\lambda_j| \leq \|A\|\text{tr}(B^\top B)$.

Proof. 1. Let U be an orthogonal matrix and let $\{\lambda_j\}_j$ be eigenvalues of A such that $U^\top AU = \text{diag}((\lambda_j)_j)$. Then we obtain

$$|(B^\top AB)_{ii}| = \left| \sum_j \lambda_j ((U^\top B)_{ji})^2 \right| \leq \|A\| \sum_j ((U^\top B)_{ji})^2 = \|A\|(B^\top B)_{ii}.$$

2. There exists an orthogonal matrix V such that $\lambda_j = (V^\top B^\top ABV)_{jj}$ for any j . Then

$$\sum_j |\lambda_j| = \sum_j |(V^\top B^\top ABV)_{jj}| \leq \|A\| \sum_j (V^\top B^\top BV)_{jj} = \|A\|\text{tr}(B^\top B)$$

by 1. \square

A.2 Proof of Lemma 4.3

Let \mathbf{A}_m be a $(k_m^1 + k_m^2) \times (k_m^1 + k_m^2)$ matrix with elements $(\mathbf{A}_m)_{ij} = 1_{i \geq j} 1_{\{i \leq k_m^1 \text{ or } j > k_m^1\}}$, $\mathbf{1}$ be a matrix with all elements equal to 1, $M_{m,*} = M_m(v_*)$ and $\hat{\mathbf{S}} = (\mathbf{A}_m^T)^{-1} M_{m,*}^{-2} \mathbf{A}_m^{-1}$.

Lemma A.8. Let $1 \leq m \leq \ell_n$, $q, q' \in \mathbb{N}$ such that $q' \geq 2q$, $A_m : \{1, \dots, k_m^1 + k_m^2\}^{q'} \rightarrow \{0, 1\}$ be a random map and $\iota : \{1, \dots, q'\} \rightarrow \{1, \dots, 2q\}$ be an injection. Assume that there exists a sequence $\{\mathcal{K}_n\}_n$ of positive numbers such that $\sum_{j_1, \dots, j_{q'}=1}^{k_m^1 + k_m^2} A_m(j_1, \dots, j_{q'}) = \bar{R}_n(\mathcal{K}_n)$. Then

$$\sum_{i_1, \dots, i_{2q}} \prod_{j=1}^q \hat{\mathbf{S}}_{i_{2j-1}, i_{2j}} A_m(i_{\iota(1)}, \dots, i_{\iota(q')}) = \bar{R}_n(k_n^{q+q'} - [q'/2] \cdot 2 \mathcal{K}_n).$$

Proof. Let $\mathbf{K}_m = \text{diag}(((k_m^1 + 1)v_{1,*})^{-1}\mathcal{E}, ((k_m^2 + 1)v_{2,*})^{-1}\mathcal{E})$. Then since $(\mathbf{A}_m M_{m,*} \mathbf{A}_m^\top)^{-1} \mathbf{1} = \mathbf{K}_m \mathbf{1}$ and $(\sum_j |((\mathbf{A}_m M_{m,*} \mathbf{A}_m^\top)^{-1})_{i,j}|) \vee (\sum_j |((\mathbf{A}_m M_{m,*} \mathbf{A}_m^\top)^{-1})_{j,i}|) \leq 2$ for any i , we obtain

$$\text{tr}(\hat{\mathbf{S}}\mathbf{1}) = \text{tr}((\mathbf{A}_m M_{m,*} \mathbf{A}_m^\top)^{-1} \mathbf{A}_m \mathbf{A}_m^\top (\mathbf{A}_m M_{m,*} \mathbf{A}_m^\top)^{-1} \mathbf{1}) = \text{tr}(\mathbf{A}_m \mathbf{A}_m^\top \mathbf{K}_m \mathbf{1} \mathbf{K}_m) = \bar{R}_n(k_n),$$

and

$$\begin{aligned} |(\hat{\mathbf{S}}\mathbf{1})^l \hat{\mathbf{S}}|_{i,j} &\leq \left| \sum_{j_1, \dots, j_{2l+2}} ((\mathbf{A}_m M_{m,*} \mathbf{A}_m^\top)^{-1})_{i,j_1} \left(\prod_{1 \leq k \leq l} (\mathbf{A}_m \mathbf{A}_m^\top)_{j_{2k-1}, j_{2k}} (\mathbf{K}_m \mathbf{1} \mathbf{K}_m)_{j_{2k}, j_{2k+1}} \right) \right. \\ &\quad \left. \times (\mathbf{A}_m \mathbf{A}_m^\top)_{j_{2l+1}, j_{2l+2}} ((\mathbf{A}_m M_{m,*} \mathbf{A}_m^\top)^{-1})_{j_{2l+2}, j} \right| \\ &= \bar{R}_n(\underline{k}_n^{-2l} \bar{k}_n^{3l+1}) = \bar{R}_n(k_n^{l+1}) \end{aligned}$$

for $l = 0, 1$.

If both i_{2j-1} and i_{2j} are outside the image of ι , we have

$$\sum_{i_{2j-1}, i_{2j}} \hat{\mathbf{S}}_{i_{2j-1}, i_{2j}} A_m(i_{\iota(1)}, \dots, i_{\iota(q')}) = \text{tr}(\hat{\mathbf{S}}\mathbf{1}) A_m(i_{\iota(1)}, \dots, i_{\iota(q')}).$$

Moreover, if both i_{2j-1} and i_{2k-1} are in the image of ι and neither i_{2j} nor i_{2k} is in it, then we have

$$\sum_{i_{2j}, i_{2k}} \hat{\mathbf{S}}_{i_{2j-1}, i_{2j}} \hat{\mathbf{S}}_{i_{2k-1}, i_{2k}} A_m(i_{\iota(1)}, \dots, i_{\iota(q')}) = (\hat{\mathbf{S}}\mathbf{1} \hat{\mathbf{S}})_{i_{2j-1}, i_{2k-1}} A_m(i_{\iota(1)}, \dots, i_{\iota(q')}).$$

Therefore there exist $\alpha_k \in \{0, 1\}$ for $1 \leq k \leq [q'/2]$, $0 \leq s \leq [(2q - q')/2]$ and a bijection $\iota' : \{1, \dots, q'\} \rightarrow \{1, \dots, q'\}$ such that $\sum_{k=1}^{[q'/2]} \alpha_k + [q'/2] + s = q - (q' - [q'/2] \cdot 2)$ and

$$\begin{aligned} &\sum_{i_1, \dots, i_{2q}} \prod_{j=1}^q \hat{\mathbf{S}}_{i_{2j-1}, i_{2j}} A_m(i_{\iota(1)}, \dots, i_{\iota(q')}) \\ &\leq \bar{R}_n(\bar{k}_n^{2(q' - [q'/2] \cdot 2)}) \sum_{j_1, \dots, j_{q'}} \prod_{k=1}^{[q'/2]} |((\hat{\mathbf{S}}\mathbf{1})^{\alpha_k} \hat{\mathbf{S}})_{j_{\iota'(2k-1)}, j_{\iota'(2k)}}| \text{tr}(\hat{\mathbf{S}}\mathbf{1})^s A_m(j_1, \dots, j_{q'}) \\ &= \bar{R}_n(k_n^{2(q' - [q'/2] \cdot 2)} \cdot k_n^{q - (q' - [q'/2] \cdot 2)} \mathcal{K}_n) = \bar{R}_n(k_n^{q+q' - [q'/2] \cdot 2} \mathcal{K}_n). \end{aligned}$$

□

Proof of Lemma 4.3.

Let $\{\tilde{\epsilon}_{i,m}\}_{1 \leq i \leq k_m^1 + k_m^2}$ and $\{\dot{\epsilon}_{i,m}\}_{1 \leq i \leq k_m^1 + k_m^2}$ be sequences of random variables defined by $\tilde{\epsilon}_{i,m} = \epsilon_{i+K_{m-1}^1+1}^{n,1}$ and $\dot{\epsilon}_{i,m} = \epsilon_{K_{m-1}^1+1}^{n,1}$ for $i \leq k_m^1$ and $\tilde{\epsilon}_{i,m} = \epsilon_{i-k_m^1+K_{m-1}^2+1}^{n,2}$ and $\dot{\epsilon}_{i,m} = \epsilon_{K_{m-1}^2+1}^{n,2}$ for $i > k_m^1$. Moreover, let $\tilde{Z}_{1,m} = (((\tilde{b}_{m,*}^\top \cdot (W_{S_i^{n,1}} - W_{S_{i-1}^{n,1}}))_{i=K_{m-1}^1+2}^{K_m^1})^\top, ((\tilde{b}_{m,*}^\top \cdot (W_{S_j^{n,2}} - W_{S_{j-1}^{n,2}}))_{j=K_{m-1}^2+2}^{K_m^2})^\top)^\top$, $\tilde{Z}_{2,m} = (((\epsilon_i^{n,1} - \epsilon_{i-1}^{n,1})_{i=K_{m-1}^1+2}^{K_m^1})^\top, ((\epsilon_j^{n,2} - \epsilon_{j-1}^{n,2})_{j=K_{m-1}^2+2}^{K_m^2})^\top)^\top$ and $\tilde{S}_{1,m,*} = \tilde{S}_{m,*} - M_{m,*}$.

Let $\tilde{U}_{1,m,*}$ be an orthogonal matrix and let $\Lambda_{1,m,*}$ be a diagonal matrix satisfying $\tilde{U}_{1,m,*} \tilde{S}_{1,m,*} \tilde{U}_{1,m,*}^\top = \Lambda_{1,m,*}$. Then since $\tilde{Z}_{1,m} | \mathcal{G}_{s_{m-1}} \sim N(0, \tilde{S}_{1,m,*})$, we have $\tilde{U}_{1,m,*} \tilde{Z}_{1,m} | \mathcal{G}_{s_{m-1}} \sim N(0, \Lambda_{1,m,*})$. Therefore, for any $q \in \mathbb{N}$ and $1 \leq j_1, \dots, j_{2q} \leq k_m^1 + k_m^2$, we obtain

$$E_m[\prod_{k=1}^{2q} (\tilde{U}_{1,m,*} \tilde{Z}_{1,m})_{j_k}] = \sum_{(l_{2k-1}, l_{2k})_{k=1}^q} \prod_{k=1}^q (\Lambda_{1,m,*})_{l_{2k-1}, l_{2k}}, \quad (\text{A.1})$$

where the summations on the right-hand side of both equations are over all q -pairs $(l_{2k-1}, l_{2k})_{k=1}^q$ of variables j_1, \dots, j_{2q} .

1. Let $\mathbf{S}'' = (\mathbf{A}_m^\top)^{-1} \mathbf{S}' \mathbf{A}_m^{-1}$, $\phi(A, B)_{i_1, \dots, i_4} = (A_{i_1, i_2} B_{i_3, i_4} + A_{i_1, i_3} B_{i_2, i_4} + A_{i_1, i_4} B_{i_2, i_3} + A_{i_2, i_3} B_{i_1, i_4} + A_{i_2, i_4} B_{i_1, i_3} + A_{i_3, i_4} B_{i_1, i_2})/2$ for square matrices A and B of the same size, and δ_{i_1, \dots, i_q} be a $\{0, 1\}$ -valued function that is equal to 1 if and only if $i_1 = \dots = i_q$. Then we have

$$\begin{aligned} & E_m[(\tilde{Z}_{2,m}^\top \mathbf{S}' \tilde{Z}_{2,m})^2] \\ &= E_m[(\mathbf{A}_m \tilde{Z}_{2,m})^\top \mathbf{S}'' (\mathbf{A}_m \tilde{Z}_{2,m})^2] = \sum_{i_1, \dots, i_4} \mathbf{S}''_{i_1, i_2} \mathbf{S}''_{i_3, i_4} E_m\left[\prod_{j=1}^4 (\tilde{\epsilon}_{i_j, m} - \dot{\epsilon}_{i_j, m})\right] \\ &= \sum_{i_1, \dots, i_4} \mathbf{S}''_{i_1, i_2} \mathbf{S}''_{i_3, i_4} \left\{ \phi(\mathbf{M}_1, \mathbf{M}_1)_{i_1, \dots, i_4} + (E[(\dot{\epsilon}_{i_1, m})^4] - 3E[(\dot{\epsilon}_{i_1, m})^2]^2) 1_{\{\max_{1 \leq j \leq 4} i_j \leq k_m^1 \text{ or } \min_{1 \leq j \leq 4} i_j > k_m^1\}} \right. \\ &\quad \left. + 2\phi(\mathbf{M}_1, \mathbf{M}_2)_{i_1, \dots, i_4} + \phi(\mathbf{M}_2, \mathbf{M}_2)_{i_1, \dots, i_4} + (E[(\tilde{\epsilon}_{i_1, m})^4] - 3E[(\tilde{\epsilon}_{i_1, m})^2]^2) \delta_{i_1, i_2, i_3, i_4} \right\}, \end{aligned}$$

where $\mathbf{M}_1 = \text{diag}(v_{1,*} \mathbf{1}, v_{2,*} \mathbf{1})$ and $\mathbf{M}_2 = \text{diag}(v_{1,*} \mathcal{E}, v_{2,*} \mathcal{E})$.

Hence, we obtain

$$\begin{aligned} E_m[(\tilde{Z}_m^\top \mathbf{S}' \tilde{Z}_m)^2] &= E_m[(\tilde{Z}_{2,m}^\top \mathbf{S}' \tilde{Z}_{2,m})^2] + \sum_{i_1, \dots, i_4} \mathbf{S}'_{i_1, i_2} \mathbf{S}'_{i_3, i_4} (\phi(\tilde{S}_{1,m,*}, \tilde{S}_{1,m,*}) + 2\phi(\tilde{S}_{1,m,*}, M_{m,*}))_{i_1, \dots, i_4} \\ &= 2\text{tr}(\tilde{S}_{m,*} \mathbf{S}' \tilde{S}_{m,*} \mathbf{S}') + \text{tr}(\tilde{S}_{m,*} \mathbf{S}')^2 + \sum_{j=1}^2 (E[(\epsilon_0^{n,j})^4] - 3v_{j,*}^2) \text{tr}(\mathbf{S}'' \mathcal{E}_{(j)} \mathbf{S}' \mathcal{E}_{(j)}) + \sum_i C_{n,i} |\mathbf{S}''_{ii}|^2, \end{aligned} \quad (\text{A.2})$$

by (A.1), where $C_{n,i} = E[(\dot{\epsilon}_{i,m})^4] - 3E[(\dot{\epsilon}_{i,m})^2]^2$ and $\mathcal{E}_{(j)}$ is a $(k_m^1 + k_m^2) \times (k_m^1 + k_m^2)$ -matrix with elements $(\mathcal{E}_{(1)})_{kl} = \delta_{k \leq k_m^1} \delta_{l \leq k_m^1}$ and $(\mathcal{E}_{(2)})_{kl} = \delta_{k > k_m^1} \delta_{l > k_m^1}$.

Lemma A.7 and the fact that $\max_i \sum_j |((\mathbf{A}_m M_{m,*} \mathbf{A}_m^\top)^{-1})_{ij}| \leq 2$ yield

$$\begin{aligned} \sum_i |\mathbf{S}''_{ii}|^2 &\leq \|M_{m,*} \mathbf{S}' M_{m,*}\| \max_i (((\mathbf{A}_m^\top)^{-1} M_{m,*}^{-2} \mathbf{A}_m^{-1})_{ii}) \cdot \|\tilde{S}_{m,*} \mathbf{S}' \tilde{S}_{m,*}\| \text{tr}((\mathbf{A}_m^\top)^{-1} \tilde{S}_{m,*}^{-2} \mathbf{A}_m^{-1}) \\ &\leq C r_n^2 \max_i \left(\sum_{j,k} (\mathbf{A}_m M_{m,*} \mathbf{A}_m^\top)^{-1}_{ij} (\mathbf{A}_m \mathbf{A}_m^\top)_{jk} (\mathbf{A}_m M_{m,*} \mathbf{A}_m^\top)^{-1}_{ki} \right) \text{tr} \left(\begin{pmatrix} M_{1,m} & 0 \\ 0 & M_{2,m} \end{pmatrix} \tilde{S}_{m,*}^{-2} \right) \\ &\leq C r_n^2 \bar{k}_n \text{tr}(\tilde{S}_{m,*}^{-1}) = \bar{R}_n(b_n^{-3/2} k_n^2) = \bar{R}_n(1). \end{aligned} \quad (\text{A.3})$$

Moreover, Lemmas A.1, A.4 and A.7 yield

$$\begin{aligned} \text{tr}(\mathbf{S}'' \mathcal{E}_{(1)} \mathbf{S}' \mathcal{E}_{(1)}) &\leq \text{tr} \left(\mathbf{S}'' \begin{pmatrix} \mathbf{1} + \mathcal{E} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{S}' \mathcal{E}_{(1)} \right) \\ &\leq C \text{tr}(\tilde{S}_{m,*}^{-1/2} \mathbf{A}_m^{-1} \mathcal{E}_{(1)} (\mathbf{A}_m^\top)^{-1} \tilde{S}_{m,*}^{-1/2}) \left\| \tilde{S}_{m,*}^{1/2} \mathbf{S}' \begin{pmatrix} M_{1,m} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{S}' \tilde{S}_{m,*}^{1/2} \right\| \\ &\leq C r_n \bar{\mathbf{\Sigma}}_n^{-1} (\tilde{S}_{m,*}^{-1})_{11} \leq C r_n \bar{\mathbf{\Sigma}}_n^{-1} (M_{m,*}^{-1})_{11} = \bar{R}_n(1), \end{aligned} \quad (\text{A.4})$$

since $(M_{m,*}^{-1})_{11} \leq v_{1,*}^{-1}$ by (4.5).

(A.2)–(A.4) and similar estimates for $\text{tr}(\mathbf{S}'' \mathcal{E}_{(2)} \mathbf{S}' \mathcal{E}_{(2)})$ yield $E_m[(\tilde{Z}_m^\top \mathbf{S}' \tilde{Z}_m)^2] = 2\text{tr}((\mathbf{S}' \tilde{S}_{m,*})^2) + \text{tr}(\mathbf{S}' \tilde{S}_{m,*})^2 + \bar{R}_n(1)$.

We next prove the estimate for $E_m[(\tilde{Z}_m^\top \mathbf{S}' \tilde{Z}_m)^q]$. Let $p \in \mathbb{N}$ satisfy $q \leq 2p$. Then it is sufficient to show that $E_m[(\tilde{Z}_m^\top \mathbf{S}' \tilde{Z}_m)^{2p}] = \bar{R}_n(b_n^{-2p} k_n^{4p})$.

Note that

$$E_m[(\tilde{Z}_{2,m}^\top M_{m,*}^{-2} \tilde{Z}_{2,m})^{2p}] = \sum_{i_1, \dots, i_{4p}} \hat{\mathbf{S}}_{i_1, i_2} \cdots \hat{\mathbf{S}}_{i_{4p-1}, i_{4p}} E_m \left[\prod_{j=1}^{4p} (\tilde{\epsilon}_{i_j, m} - \dot{\epsilon}_{i_j, m}) \right],$$

and there exist $\{0, 1\}$ -valued maps $\{A_{l,m}\}_l$, constants C_l , positive integers $\{q'_l\}$ not greater than $4p$ and injections $\{\iota_l\}$ such that $E_m[\prod_{j=1}^{4p}(\tilde{\epsilon}_{i_j,m} - \dot{\epsilon}_{i_j,m})] = \sum_l C_l A_{l,m}(i_{\iota_l(1)}, \dots, i_{\iota_l(q'_l)})$ and $\sum_{j_1, \dots, j_{q'_l}} A_{l,m}(j_1, \dots, j_{q'_l}) = \bar{R}_n(k_n^{[q'_l/2]})$ for any l . Then Lemma A.8 yields

$$E_m[(\tilde{Z}_{2,m}^\top M_{m,*}^{-2} \tilde{Z}_{2,m})^{2p}] = \bar{R}_n(k_n^{4p}), \quad (\text{A.5})$$

and therefore Lemma A.7 yields

$$E_m[(\tilde{Z}_{2,m}^\top \mathbf{S}' \tilde{Z}_{2,m})^{2p}] \leq \|M_{m,*} \mathbf{S}' M_{m,*}\|^{2p} E_m[(\tilde{Z}_{2,m}^\top M_{m,*}^{-2} \tilde{Z}_{2,m})^{2p}] = \bar{R}_n(b_n^{-2p} k_n^{4p}).$$

Moreover, (A.1) yields

$$E_m[(\tilde{Z}_{1,m}^\top \mathbf{S}' \tilde{Z}_{1,m})^{2p}] \leq C_p \sum_{\substack{\gamma=(\gamma_1, \dots, \gamma_L); \\ L \in \mathbb{N}, \gamma_k \geq 1, \sum_k \gamma_k = 2p}} \prod_k \text{tr}((\mathbf{S}' \tilde{S}_{1,m,*})^{\gamma_k}) = \bar{R}_n(b_n^{-p} k_n^{2p}).$$

Furthermore, by calculating the expectation of $\tilde{Z}_{1,m}$ and using (A.1) and Lemma A.7, we have

$$\begin{aligned} E_m[(\tilde{Z}_{1,m}^\top \mathbf{S}' \tilde{Z}_{2,m})^{2p}] &= \left(\sum_{(l_{2k-1}, l_{2k})_{k=1}^q} 1 \right) E_m[(\tilde{Z}_{2,m}^\top \mathbf{S}' \tilde{S}_{1,m,*} \mathbf{S}' \tilde{Z}_{2,m})^p] \\ &\leq (2p-1)!! \|M_{m,*} \mathbf{S}' \tilde{S}_{1,m,*} \mathbf{S}' M_{m,*}\|^p E_m[(\tilde{Z}_{2,m}^\top M_{m,*}^{-2} \tilde{Z}_{2,m})^p] = \bar{R}_n(b_n^{-p} k_n^{2p}). \end{aligned}$$

Then we obtain $E_m[(\tilde{Z}_m^\top \mathbf{S}' \tilde{Z}_m)^{2p}] = \bar{R}_n(b_n^{-2p} k_n^{4p})$.

For the estimate of $E_m[(\tilde{Z}_m^\top \mathbf{S}' \tilde{Z}_m)^4]$, we have $E_m[(\tilde{Z}_{1,m}^\top \mathbf{S}' \tilde{Z}_{1,m})^4] = \bar{R}_n(b_n^{-2} k_n^4)$ and $E_m[(\tilde{Z}_{1,m}^\top \mathbf{S}' \tilde{Z}_{2,m})^4] = \bar{R}_n(b_n^{-2} k_n^4)$ by the above results. Moreover, we have

$$E_m[(\tilde{Z}_{2,m}^\top \mathbf{S}' \tilde{Z}_{2,m})^4] = \sum_{i_1, \dots, i_8} \prod_{k=1}^4 ((\mathbf{A}_m^\top)^{-1} \mathbf{S}' \mathbf{A}_m^{-1})_{i_{2k-1}, i_{2k}} E_m[\prod_{k=1}^8 (\tilde{\epsilon}_{i_k,m} - \dot{\epsilon}_{i_k,m})] \quad (\text{A.6})$$

and there exist $\{0, 1\}$ -valued maps $\{A'_l\}_l$, constants C'_l , positive integers $\{q''_l\}$ not greater than 8 and injections $\{\iota'_l\}$ such that $\sum_{j_1, \dots, j_{q''_l}} A'_l(j_1, \dots, j_{q''_l}) = \bar{R}_n(k_n^{[q''_l/2] \wedge 3})$ and

$$E_m[\prod_{j=1}^8 (\tilde{\epsilon}_{i_j,m} - \dot{\epsilon}_{i_j,m})] = \sum_{(l_{2k-1}, l_{2k})_{k=1}^4} \prod_{k=1}^4 \delta_{l_{2k-1}, l_{2k}} + \sum_l C'_l A'_l(i_{\iota'_l(1)}, \dots, i_{\iota'_l(q''_l)}), \quad (\text{A.7})$$

where the summation in the first term of the right-hand side is over all 4-pairs $(l_{2k-1}, l_{2k})_{k=1}^4$ of variables i_1, \dots, i_8 .

Let $\tilde{\mathbf{A}}_m = M_{m,*} \mathbf{A}_m^\top$, then a simple calculation shows that

$$(\tilde{\mathbf{A}}_m^{-1})_{i,j} = \begin{cases} (k_m^{\mathbf{k}(i)} + 1)^{-1} (j - k_m^1 1_{\{\mathbf{k}(i)=2\}}) v_{\mathbf{k}(i),*}^{-1}, & \mathbf{k}(i) = \mathbf{k}(j) \text{ and } i \geq j, \\ -(k_m^{\mathbf{k}(i)} + 1)^{-1} (k_m^{\mathbf{k}(i)} - j + 1 + k_m^1 1_{\{\mathbf{k}(i)=2\}}) v_{\mathbf{k}(i),*}^{-1}, & \mathbf{k}(i) = \mathbf{k}(j) \text{ and } i < j, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, we have

$$\begin{aligned} |((\tilde{\mathbf{A}}_m^\top)^{-1} \mathbf{S}' (\tilde{\mathbf{A}}_m^{-1})_{ij})| &= |(\tilde{\mathbf{A}}_m^{-1} M_{m,*} \mathbf{S}' M_{m,*} (\tilde{\mathbf{A}}_m^\top)^{-1})_{ij}| \\ &\leq \sum_{k_1, k_2} |(\tilde{\mathbf{A}}_m^{-1})_{i,k_1} (M_{m,*} \mathbf{S}' M_{m,*})_{k_1 k_2} (\tilde{\mathbf{A}}_m^{-1})_{j,k_2}| \\ &\leq \left(\sum_k ((\tilde{\mathbf{A}}_m^{-1})_{i,k})^2 \right)^{1/2} \|M_{m,*} \mathbf{S}' M_{m,*}\| \left(\sum_k ((\tilde{\mathbf{A}}_m^{-1})_{j,k})^2 \right)^{1/2} = \bar{R}_n(b_n^{-1} k_n). \quad (\text{A.8}) \end{aligned}$$

Similarly, we have

$$|((\mathbf{A}_m^\top)^{-1} \mathbf{S}' \mathbf{A}_m^{-1} \mathbf{1} (\mathbf{A}_m^\top)^{-1} \mathbf{S}' \mathbf{A}_m^{-1})_{ij}| = \bar{R}_n(b_n^{-2} k_n^2). \quad (\text{A.9})$$

Then (A.6)–(A.9) and a similar argument to the proof of Lemma A.8 yield

$$\begin{aligned} E_m[(\tilde{Z}_{2,m}^\top \mathbf{S}' \tilde{Z}_{2,m})^4] &= \sum_{i_1, \dots, i_8} \prod_{k=1}^4 ((\mathbf{A}_m^{-1})^\top \mathbf{S}' \mathbf{A}_m^{-1})_{i_{2k-1}, i_{2k}} \sum_{(l_{2k-1}, l_{2k})_{k=1}^4} \prod_{k=1}^4 \delta_{l_{2k-1}, l_{2k}} + \bar{R}_n((b_n^{-4} k_n^7) \vee (b_n^{-2} k_n^4)) \\ &= \bar{R}_n((b_n^{-4} k_n^7) \vee (b_n^{-2} k_n^4)). \end{aligned}$$

2. Let $\{\tilde{I}_{i,m}\}_{i=1}^{k_m^1+k_m^2}$ and $\{\mathbf{k}(i)\}_{i=1}^{k_m^1+k_m^2}$ be defined by $\tilde{I}_{i,m} = I_{i,m}^1$, $\mathbf{k}(i) = 1$ for $1 \leq i \leq k_m^1$ and $\tilde{I}_{i,m} = I_{i-k_m^1,m}^2$, $\mathbf{k}(i) = 2$ for $k_m^1 < i \leq k_m^1 + k_m^2$. Since $(Z_m - \tilde{Z}_{2,m} \pm \tilde{Z}_{1,m})_i = \int_{\tilde{I}_{i,m}} (b_{t,*}^{\mathbf{k}(i)} \pm \tilde{b}_{m,*}^{\mathbf{k}(i)}) dW_t + \mu_{s_{m-1}}^{\mathbf{k}(i)} |\tilde{I}_{i,m}| + \int_{\tilde{I}_{i,m}} (\mu_t^{\mathbf{k}(i)} - \mu_{s_{m-1}}^{\mathbf{k}(i)}) dt$, we have $(Z_m - \tilde{Z}_m)^\top \mathbf{S}' (Z_m + \tilde{Z}_m) = \Psi_{m,1} + \Psi_{m,2} + \Psi_{m,3}$, where

$$\begin{aligned} \Psi_{m,1} &= 2(Z_m - \tilde{Z}_m)^\top \mathbf{S}' \tilde{Z}_{2,m} + \sum_{k=1}^2 \sum_{i,j} \mathbf{S}'_{i,j} \int_{\tilde{I}_{i,m}} (b_{t,*}^{\mathbf{k}(i)} + (-1)^k \tilde{b}_{m,*}^{\mathbf{k}(i)}) \int_{\tilde{I}_{j,m} \cap [0,t)} (b_{s,*}^{\mathbf{k}(j)} + (-1)^{k-1} \tilde{b}_{m,*}^{\mathbf{k}(j)}) dW_s dW_t \\ &\quad + \sum_{k=1}^2 \sum_{i,j} \mathbf{S}'_{i,j} \mu_{s_{m-1}}^{\mathbf{k}(i)} |\tilde{I}_{i,m}| \int_{\tilde{I}_{j,m}} (b_{t,*}^{\mathbf{k}(j)} + (-1)^k \tilde{b}_{m,*}^{\mathbf{k}(j)}) dW_t, \end{aligned}$$

$$E_m[\Psi_{m,2}] = 0, \Psi_{m,2} = \bar{R}_n(b_n^{-1} k_n^{3/2}) \text{ and } \Psi_{m,3} = \bar{R}_n(b_n^{-3/2} k_n^2).$$

Then the Burkholder–Davis–Gundy inequality yields

$$\begin{aligned} E_\Pi \left[\left(\sum_m (Z_m - \tilde{Z}_m)^\top \mathbf{S}' (Z_m + \tilde{Z}_m) \right)^q \right] &\leq C E_\Pi \left[\left(\sum_m (\Psi_{m,1} + \Psi_{m,2})^2 \right)^{\frac{q}{2}} \right] + \bar{R}_n(b_n^{-\frac{q}{2}} k_n^q) \\ &\leq C E_\Pi \left[\left(\sum_m E_m[\Psi_{m,1}^2] \right)^{\frac{q}{2}} \right] + C E_\Pi \left[\left(\sum_m \bar{E}_m[\Psi_{m,1}^2] \right)^{\frac{q}{2}} \right] + \bar{R}_n(b_n^{-\frac{q}{2}} k_n^q). \end{aligned}$$

We can rewrite $(Z_{1,m} - \tilde{Z}_{1,m})_i = \mathbf{L}_i^1 + \mathbf{L}_i^2 + \mathbf{L}_i^3 + \bar{R}_n((r_n^{1/2} \ell_n^{-3/2}) \vee r_n)$, where $\mathbf{L}_i^1 = \sum_j \xi_j^1 \int_{\tilde{I}_{i,m}} (t - s_{m-1}) dW_t^j$, $\mathbf{L}_i^2 = \sum_{j,k} \xi_{j,k}^2 \int_{\tilde{I}_{i,m}} (W_t^k - W_{s_{m-1}}^k) dW_t^j$ and $\mathbf{L}_i^3 = \sum_{j,k,l} \xi_{j,k,l}^3 \int_{\tilde{I}_{i,m}} \int_{s_{m-1}}^t (W_s^l - W_{s_{m-1}}^l) dW_s^k dW_t^j$ for some $\mathcal{G}_{s_{m-1}}$ -measurable random variables ξ_j^1 , $\xi_{j,k}^2$, and $\xi_{j,k,l}^3$ with bounded moments. Let $\mathbf{L}^j = (\mathbf{L}_i^j)_i$. Then, for any $p \in \mathbb{N}$, Lemma A.7 and (A.5) yield

$$\begin{aligned} E_m[\Psi_{m,1}^{2p}] &\leq C E_m[(\tilde{Z}_{2,m}^\top \mathbf{S}' (Z_m - \tilde{Z}_m) (Z_m - \tilde{Z}_m)^\top \mathbf{S}' \tilde{Z}_{2,m})^p] + C E_m[((Z_{1,m} + \tilde{Z}_{1,m})^\top \mathbf{S}' (Z_{1,m} - \tilde{Z}_{1,m}))^{2p}] \\ &\quad + \bar{R}_n((b_n^{-1} k_n^{3/2})^{2p}) \\ &\leq C E_m[\|M_{m,*} \mathbf{S}' (Z_m - \tilde{Z}_m) (Z_m - \tilde{Z}_m)^\top \mathbf{S}' M_{m,*}\|^p (\tilde{Z}_{2,m}^\top M_{m,*}^{-2} \tilde{Z}_{2,m})^p] + C \sum_{j=1}^3 E_m[(\tilde{Z}_{1,m}^\top \mathbf{S}' \mathbf{L}^j)^{2p}] \\ &\quad + C E_m[(\mathbf{L}^2)^\top \mathbf{S}' \mathbf{L}^2]^{2p} + \bar{R}_n((b_n k_n r_n (\ell_n^{-3/2} \vee r_n^{1/2}))^{2p}) + \bar{R}_n((b_n^{-1} k_n^{3/2})^{2p}) \\ &= C \sum_{j=1}^3 E_m[(\tilde{Z}_{1,m}^\top \mathbf{S}' \mathbf{L}^j)^{2p}] + C E_m[(\mathbf{L}^2)^\top \mathbf{S}' \mathbf{L}^2]^{2p} + \bar{R}_n(b_n^{-2p} k_n^{4p}) + \bar{R}_n(b_n^{-3p} k_n^{5p}). \end{aligned} \quad (\text{A.10})$$

Moreover, we can see that there exists a positive constant C_p such that

$$\begin{aligned} &E_m \left[\sum_{\substack{i_1, \dots, i_{2p} \\ j_1, \dots, j_{2p}}} \prod_{k=1}^{2p} (W^{p1,k}(\tilde{I}_{i_k,m}) \int_{\tilde{I}_{j_k,m}} (W_t^{p2,k} - W_{s_{m-1}}^{p2,k}) dW_t^{p3,k}) \right] \\ &\leq C_p \sum_{\substack{i_1, \dots, i_{2p} \\ j_1, \dots, j_{2p}}} \sum_{(l_{2q-1}, l_{2q})_{q=1}^{2p-\alpha}, \alpha} \left(\prod_{q=1}^{2p-\alpha} |\tilde{I}_{l_{2q-1},m} \cap \tilde{I}_{l_{2q},m}| \right) r_n^{2\alpha} (s_m - s_{m-1})^{-p+\alpha} \end{aligned}$$

for any $\{p_{l,k}\}_{1 \leq l \leq 3, 1 \leq k \leq 2p} \subset \{1, 2\}$, where the summation in the right-hand side is taken over $0 \leq \alpha \leq p$ and $(2p - \alpha)$ disjoint pairs $(l_{2q-1}, l_{2q})_{q=1}^{2p-\alpha}$ in the variables $i_1, \dots, i_{2p}, j_1, \dots, j_{2p}$. Here we used the fact that all $6p$ factors $(W^{p_{1,k}}(\tilde{I}_{i_k, m}), \int_{\tilde{I}_{j_k, m}} \cdot dW_t^{p_{3,k}}, (W_t^{p_{2,k}} - W_{s_{m-1}}^{p_{2,k}})_{t \in \tilde{I}_{j_k, m}})_{k=1}^{2p}$ should be separated into $3p$ pairs in the non-zero terms. 2α represents the number of pairs with the form $(W^{p_{1,k}}(\tilde{I}_{i_k, m}), (W_t^{p_{2,k'}} - W_{s_{m-1}}^{p_{2,k'}})_{t \in \tilde{I}_{j_{k'}, m}})$ or $(\int_{\tilde{I}_{j_k, m}} \cdot dW_t^{p_{3,k}}, (W_t^{p_{2,k'}} - W_{s_{m-1}}^{p_{2,k'}})_{t \in \tilde{I}_{j_{k'}, m}})$. Therefore we obtain

$$E_m[(\tilde{Z}_{1,m}^\top \mathbf{S}' \mathbf{L}^2)^{2p}] = \bar{R}_n((b_n^{1/2} \ell_n^{-1})^{2p-\alpha} (b_n^{3/2} \ell_n^{-1} k_n)^\alpha r_n^{2\alpha} \ell_n^{-p+\alpha}) = \bar{R}_n((b_n^{-1} k_n^{3/2})^{2p}). \quad (\text{A.11})$$

Similar arguments for $E_m[(\tilde{Z}_{1,m}^\top \mathbf{S}' \mathbf{L}^1)^{2p}]$, $E_m[(\tilde{Z}_{1,m}^\top \mathbf{S}' \mathbf{L}^3)^{2p}]$ and $E_m[(\mathbf{L}^2)^\top \mathbf{S}' \mathbf{L}^2]^{2p}$ yield $E_m[\Psi_{m,1}^{2p}] = \bar{R}_n(((b_n^{-1} k_n^2) \vee (b_n^{-3/2} k_n^{5/2}))^{2p})$, and consequently we obtain $E_\Pi[(\sum_m \Psi_{m,1}^4)^{q/4}] = \bar{R}_n((b_n^{-3} k_n^7)^{q/4})$.

Furthermore, since $b_{t,*}^{\mathbf{k}(i)} + \tilde{b}_{m,*}^{\mathbf{k}(i)} = 2\tilde{b}_{m,*}^{\mathbf{k}(i)} + (b_{t,*}^{\mathbf{k}(i)} - \tilde{b}_{m,*}^{\mathbf{k}(i)})$ and $b_{t,*}^{\mathbf{k}(i)} - \tilde{b}_{m,*}^{\mathbf{k}(i)} = \sum_j \tilde{\xi}_j^1 (W_t^j - W_{s_{m-1}}^j) + \tilde{\xi}^2 (t - s_{m-1}) + \sum_{j,k} \tilde{\xi}_{j,k}^3 \int_{s_{m-1}}^t (W_s^j - W_{s_{m-1}}^j) dW_s^k + \bar{R}_n(\ell_n^{-3/2})$ for some $\mathcal{G}_{s_{m-1}}$ -measurable random variables $\tilde{\xi}_j^1$, $\tilde{\xi}^2$ and $\tilde{\xi}_{j,k}^3$ with bounded moments, an argument similar to the one above and Lemmas A.2 and A.7 yield

$$\begin{aligned} E_m[\Psi_{m,1}^2] &\leq C \|\tilde{S}_{m,*}^{1/2} \mathbf{S}' M_{m,*} \mathbf{S}' \tilde{S}_{m,*}^{1/2}\| E_m[(Z_m - \tilde{Z}_m)^\top \tilde{S}_{m,*}^{-1} (Z_m - \tilde{Z}_m)] + \bar{R}_n(b_n k_n r_n \ell_n^{-2}) \\ &\quad + C \sum_{i_1, i_2, j_1, j_2} |\mathbf{S}'_{i_1, j_1}| |\mathbf{S}'_{i_2, j_2}| |\tilde{I}_{i_1, m} \cap \tilde{I}_{i_2, m}| (\ell_n^{-1} |\tilde{I}_{j_1, m} \cap \tilde{I}_{j_2, m}| + r_n^2 + r_n^{3/2} \ell_n^{-3/2}) \\ &\quad + \left\| \mathbf{S}' \left\{ \sum_{k_1, k_2=1}^2 \int_{\tilde{I}_{i,m} \cap \tilde{I}_{j,m}} E_m[(b_{t,*}^{\mathbf{k}(i)} + (-1)^{k_1} \tilde{b}_{m,*}^{\mathbf{k}(i)})(b_{t,*}^{\mathbf{k}(i)} + (-1)^{k_2} \tilde{b}_{m,*}^{\mathbf{k}(i)})] dt \right\}_{i,j} \mathbf{S}' \right\| \bar{R}_n(k_n r_n^2) \\ &= \bar{R}_n(b_n^{-1} k_n^{3/2}) + \bar{R}_n(b_n^{-2} k_n^3) + \bar{R}_n((b_n^{3/2} \ell_n^{-1})^2 r_n (\ell_n^{-1} r_n + k_n r_n^{3/2} \ell_n^{-3/2})) + \bar{R}_n(b_n^{-1} k_n) \\ &= \bar{R}_n((b_n^{-2} k_n^3) \vee (b_n^{-3} k_n^{9/2})). \end{aligned}$$

Therefore, we have $E_\Pi[(\sum_m E_m[\Psi_{m,1}^2])^{q/2}] = \bar{R}_n(((b_n^{-1} k_n^2) \vee (b_n^{-2} k_n^{7/2}))^{q/2})$, which completes the proof of point 2.

Then point 3 is easily obtained by the proof of point 2 since we only need the estimate for $E_\Pi[(\sum_m \Psi_{m,1}^2)]$ if $q = 2$. \square

A.3 An additional lemma

Lemma A.9. *Let e_n be a sequence of positive numbers, \mathcal{S} be an open set in a Euclidean space, $A_n(\lambda)$ and $B_n(\lambda)$ be sequences of positive-valued random variables, and $C_n(\lambda)$ be a sequence of non-negative-valued random variables for $\lambda \in \mathcal{S}$. Assume that $A_n(\lambda)$, $B_n(\lambda)$, and $C_n(\lambda)$ are C^3 with respect to λ , $e_n \rightarrow \infty$, $\sup_{0 \leq k \leq 3, \lambda \in \mathcal{S}} (|\partial_\lambda^k A_n| \vee |\partial_\lambda^k B_n|) = O_p(e_n^{-1})$, and $\sup_{0 \leq k \leq 3, \lambda \in \mathcal{S}} |\partial_\lambda^k C_n| = O_p(e_n^{-2})$ as $n \rightarrow \infty$, $C_n < A_n B_n$ a.s. for any $n \in \mathbb{N}$, and $\lim_{n \rightarrow \infty} (e_n^2 (A_n B_n - C_n)) > 0$ a.s. Then*

$$\sup_{\lambda \in \mathcal{S}} \left| \partial_\lambda^k \left(\sum_{p=1}^\infty \int_0^\pi \frac{C_n^p}{f_p(A_n, x) f_p(B_n, x)} dx - \frac{\pi C_n}{\sqrt{2} P_n \sqrt{A_n B_n - C_n}} \right) \right| = O_p(e_n^{-\frac{3}{2}}), \quad (\text{A.12})$$

$$\sup_{\lambda \in \mathcal{S}} \left| \partial_\lambda^k \left(\sum_{p=1}^\infty \int_0^\pi \frac{C_n^p f_1(A_n, x)}{f_p(A_n, x) f_p(B_n, x)} dx - \frac{\pi C_n (A_n + \sqrt{A_n B_n - C_n})}{\sqrt{2} P_n \sqrt{A_n B_n - C_n}} \right) \right| = O_p(e_n^{-\frac{5}{2}}), \quad (\text{A.13})$$

$$\sup_{\lambda \in \mathcal{S}} \left| \partial_\lambda^k \left(\sum_{p=1}^\infty \frac{1}{p} \int_0^\pi \frac{C_n^p}{f_p(A_n, x) f_p(B_n, x)} dx - \pi (\sqrt{A_n} + \sqrt{B_n}) + \frac{\pi}{\sqrt{2}} P_n \right) \right| = O_p(e_n^{-1}) \quad (\text{A.14})$$

for $0 \leq k \leq 3$, where $f_p(a, x) = (a + 2(1 - \cos x))^p$ and

$$P_n = \sqrt{A_n + B_n + \sqrt{(A_n - B_n)^2 + 4C_n}} + \sqrt{A_n + B_n - \sqrt{(A_n - B_n)^2 + 4C_n}}.$$

Proof. An elementary calculation yields

$$\begin{aligned}
& \sum_{p=1}^{\infty} \int_0^{\pi} \frac{C_n^p}{(A_n + 2(1 - \cos x))^p (B_n + 2(1 - \cos x))^p} dx \\
&= \int_{-\infty}^{\infty} \left(1 - \frac{C_n}{(A_n + \frac{4t^2}{1+t^2})(B_n + \frac{4t^2}{1+t^2})} \right)^{-1} \frac{C_n}{(A_n + \frac{4t^2}{1+t^2})(B_n + \frac{4t^2}{1+t^2})} \frac{1}{1+t^2} dt \\
&= \int_{-\infty}^{\infty} \frac{C_n}{(A_n + \frac{4t^2}{1+t^2})(B_n + \frac{4t^2}{1+t^2}) - C_n} \frac{1}{1+t^2} dt \\
&= \int_{-\infty}^{\infty} \frac{C_n(1+t^2)}{((A_n+4)t^2 + A_n)((B_n+4)t^2 + B_n) - C_n(1+t^2)^2} dt \\
&= \int_{-\infty}^{\infty} \frac{C_n(1+t^2)}{(A_n B_n + 4A_n + 4B_n - C_n + 16)t^4 + 2(A_n B_n + 2A_n + 2B_n - C_n)t^2 + A_n B_n - C_n} dt. \quad (\text{A.15})
\end{aligned}$$

We only consider the case $(A_n - B_n)^2 + 16C_n > 0$ a.s. We can easily obtain the results for the other case with a slight modification.

Let

$$\alpha_1 = \frac{-2A_n - 2B_n - \sqrt{4(A_n - B_n)^2 + 16C_n}}{16} \quad \text{and} \quad \alpha_2 = \frac{-2A_n - 2B_n + \sqrt{4(A_n - B_n)^2 + 16C_n}}{16},$$

then we have $\alpha_1 < \alpha_2 < 0$ by the assumptions. Moreover, we can calculate the right-hand side of (A.15) as

$$\begin{aligned}
& (1 + O_p(e_n^{-1})) \int_{-\infty}^{\infty} \frac{C_n(1+t^2)}{16(t^2 - \alpha_1)(t^2 - \alpha_2)} dt \\
&= (1 + O_p(e_n^{-1})) \frac{2\pi i}{16} \left[\frac{C_n(1+\alpha_1)}{2\sqrt{-\alpha_1}i(\alpha_1 - \alpha_2)} + \frac{C_n(1+\alpha_2)}{2\sqrt{-\alpha_2}i(\alpha_2 - \alpha_1)} \right] \\
&= (1 + O_p(e_n^{-1})) \frac{\pi}{16} \frac{C_n(\sqrt{-\alpha_2} - \sqrt{-\alpha_1})(1 + \sqrt{\alpha_1\alpha_2})}{\sqrt{\alpha_1\alpha_2}(\alpha_1 - \alpha_2)} = \frac{\pi}{16} \frac{C_n}{\sqrt{\alpha_1\alpha_2}(\sqrt{-\alpha_1} + \sqrt{-\alpha_2})} + O_p(e_n^{-3/2}).
\end{aligned}$$

Therefore, we obtain (A.12) with $k = 0$ by noting that $16\alpha_1\alpha_2 = (A_n B_n - C_n)(1 + O_p(e_n^{-1}))$.

We similarly have (A.12) with $1 \leq k \leq 3$ and (A.13) with $0 \leq k \leq 3$.

Furthermore, we obtain $\int_{-\infty}^{\infty} (1+t^2)^{-1} \log(t^2 + \alpha^2) dt = 2\pi \log(\alpha + 1)$ for any $\alpha > 0$ by the residue theorem.

Therefore, we obtain

$$\begin{aligned}
& \partial_{\lambda}^k \sum_{p=1}^{\infty} \frac{1}{p} \int_0^{\pi} \frac{C_n^p}{f_p(A_n, x) f_p(B_n, x)} dx \\
&= -\partial_{\lambda}^k \int_{-\infty}^{\infty} \frac{1}{1+t^2} \log \left(1 - \frac{C_n}{(A_n + 4t^2/(1+t^2))(B_n + 4t^2/(1+t^2))} \right) dt \\
&= -\partial_{\lambda}^k \int_{-\infty}^{\infty} \frac{1}{1+t^2} \log \left(\frac{(t^2 - \alpha_1)(t^2 - \alpha_2)}{(t^2 + A_n/(A_n + 4))(t^2 + B_n/(B_n + 4))} \right) dt + O_p(e_n^{-1}) \\
&= -2\pi \partial_{\lambda}^k \left(\log(1 + \sqrt{-\alpha_1}) + \log(1 + \sqrt{-\alpha_2}) - \log \left(1 + \sqrt{\frac{A_n}{A_n + 4}} \right) - \log \left(1 + \sqrt{\frac{B_n}{B_n + 4}} \right) \right) + O_p(e_n^{-1}) \\
&= 2\pi \partial_{\lambda}^k (\sqrt{A_n}/2 + \sqrt{B_n}/2 - \sqrt{-\alpha_1} - \sqrt{-\alpha_2}) + O_p(e_n^{-1}),
\end{aligned}$$

which completes the proof. \square

Acknowledgement

The author is grateful to Yuta Koike for very useful discussions on numerical analysis of quadratic covariation estimators. He had a great contribution on constructing the ‘cce’ function in the ‘yuima’ R package as well. This work was supported by Japan Society for the Promotion of Science KAKENHI Grant Number 15K21598.

References

- [1] R. A. Adams and J. J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.
- [2] Y. Aït-Sahalia, J. Fan, and D. Xiu. High-frequency covariance estimates with noisy and asynchronous financial data. *J. Amer. Statist. Assoc.*, 105(492):1504–1517, 2010.
- [3] O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536, 2008.
- [4] O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics*, 162(2):149–169, 2011.
- [5] M. Bibinger, N. Hautsch, P. Malec, and M. Reiss. Estimating the quadratic covariation matrix from noisy observations: local method of moments and efficiency. *Ann. Statist.*, 42(4):80–114, 2014.
- [6] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [7] K. Christensen, S. Kinnebrock, and M. Podolskij. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *J. Econometrics*, 159(1):116–133, 2010.
- [8] K. Christensen, M. Podolskij, and M. Vetter. On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes. *J. Multivariate Anal.*, 120:59–84, 2013.
- [9] J. C. Cox, J. E. Ingersoll, Jr., and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.
- [10] P. Doukhan and S. Louhichi. A new weak dependence condition and applications to moment inequalities. *Stochastic Process. Appl.*, 84(2):313–342, 1999.
- [11] V. Genon-Catalot and J. Jacod. Estimation of the diffusion coefficient for diffusion processes: random sampling. *Scand. J. Statist.*, 21(3):193–221, 1994.
- [12] A. Gloter and J. Jacod. Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM Probab. Statist.*, 5:225–242 (electronic), 2001.
- [13] A. Gloter and J. Jacod. Diffusions with measurement errors. II. Optimal estimators. *ESAIM Probab. Statist.*, 5:243–260 (electronic), 2001.
- [14] E. Gobet. Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. *Bernoulli*, 7(6):899–912, 2001.
- [15] T. Hayashi and N. Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379, 2005.
- [16] T. Hayashi and N. Yoshida. Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes. *Ann. Inst. Statist. Math.*, 60(2):367–406, 2008.
- [17] T. Hayashi and N. Yoshida. Nonsynchronous covariation process and limit theorems. *Stochastic Process. Appl.*, 121(10):2416–2454, 2011.
- [18] J. Jacod. *Calcul stochastique et problèmes de martingales*, volume 714 of *Lecture Notes in Mathematics*. Springer, Berlin, 1979.
- [19] J. Jacod. *On continuous conditional Gaussian martingales and stable convergence in law*, volume 1655 of *Lecture Notes in Math.* Springer, Berlin, 1997.
- [20] J. Jacod, Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter. Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Process. Appl.*, 119(7):2249–2276, 2009.
- [21] P. Jeganathan. On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā Ser. A*, 44(2):173–212, 1982.
- [22] P. Jeganathan. Some asymptotic properties of risk functions when the limit of the experiment is mixed normal. *Sankhyā Ser. A*, 45(1):66–87, 1983.
- [23] P. Malliavin and M. E. Mancino. Fourier series method for measurement of multivariate volatilities. *Finance Stoch.*, 6(1):49–61, 2002.
- [24] P. Malliavin and M. E. Mancino. A Fourier transform method for nonparametric estimation of multivariate volatility. *Ann. Statist.*, 37(4):1983–2010, 2009.

- [25] T. Ogihara. Local asymptotic mixed normality property for nonsynchronously observed diffusion processes. *Bernoulli*, in press, 2014.
- [26] T. Ogihara and N. Yoshida. Quasi-likelihood analysis for nonsynchronously observed diffusion processes. *Stochastic Process. Appl.*, 124(9):2954–3008, 2014.
- [27] M. Podolskij and M. Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15(3):634–658, 2009.
- [28] M. Uchida. Contrast-based information criterion for ergodic diffusion processes from discrete observations. *Ann. Inst. Statist. Math.*, 62(1):161–187, 2010.
- [29] M. Uchida and N. Yoshida. Quasi likelihood analysis of volatility and nondegeneracy of statistical random field. *Stochastic Process. Appl.*, 123(7):2851–2876, 2013.
- [30] N. Yoshida. Polynomial type large deviation inequalities and convergence of statistical random fields. *ISM Research Memorandum 1021*, 2006.
- [31] N. Yoshida. Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann. Inst. Statist. Math.*, 63(3):431–479, 2011.
- [32] L. Zhang, P. A. Mykland, and Y. Aït-Sahalia. A tale of two time scales: determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 100(472):1394–1411, 2005.